



**UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE QUÍMICA**

**BRUNO CÉSAR DO AMARAL**

**UMA NOVA METODOLOGIA PARA LIGAÇÃO CRUZADA ENTRE RESÍDUOS DE  
AMINOÁCIDOS NUCLEOFÍLICOS**

**CAMPINAS  
2017**

**BRUNO CÉSAR DO AMARAL**

**UMA NOVA METODOLOGIA PARA LIGAÇÃO CRUZADA ENTRE  
RESÍDUOS DE AMINOÁCIDOS NUCLEOFÍLICOS**

**Dissertação de Mestrado apresentada ao Instituto de  
Química da Universidade Estadual de Campinas como  
parte dos requisitos exigidos para a obtenção do título  
de Mestre em Química na área de Química Orgânica**

**Orientador: Prof. Dr. Fabio Cesar Gozzo**

**ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO  
DEFENDIDA PELO ALUNO BRUNO CÉSAR DO AMARAL, E ORIENTADA PELO  
PROF. DR. FABIO CESAR GOZZO.**

**CAMPINAS  
2017**

**Agência(s) de fomento e nº(s) de processo(s):** CNPq, 134660/2015-3

**ORCID:** <http://orcid.org/http://orcid.org/00>

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Química  
Camila Barleta Fullin - CRB 8462

Am13n      Amaral, Bruno César do, 1991-  
Uma nova metodologia para ligação cruzada entre resíduos de  
aminoácidos nucleofílicos / Bruno César do Amaral. – Campinas, SP : [s.n.],  
2017.

Orientador: Fabio Cesar Gozzo.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de  
Química.

1. Espectrometria de massas. 2. Ligação cruzada. 3. Proteômica  
estrutural. 4. Proteínas. 5. Proteínas - Estrutura. I. Gozzo, Fabio Cesar, 1972-.  
II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** A novel methodology for cross-linking targeting nucleophilic aminoacid residues

**Palavras-chave em inglês:**

Mass spectrometry

Cross-linking

Structural proteomics

Proteins

Proteins - Structure

**Área de concentração:** Química Orgânica

**Titulação:** Mestre em Química na área de Química Orgânica

**Banca examinadora:**

Fábio Cesar Gozzo

Giuseppe Palmisano

Carlos Roque Duarte Correia

**Data de defesa:** 25-07-2017

**Programa de Pós-Graduação:** Química

## **BANCA EXAMINADORA**

Prof. Dr. Fabio Cesar Gozzo (Orientador)

Prof. Dr. Giuseppe Palmisano (ICB-USP)

Prof. Dr. Carlos Roque Duarte Correia (IQ-UNICAMP)

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no processo de vida acadêmica do(a) aluno(a).

Este exemplar corresponde à redação final da Dissertação de Mestrado defendida pelo aluno **BRUNO CÉSAR DO AMARAL**, aprovada pela Comissão Julgadora em 25 de julho de 2017.

*“Tenha em mente que tudo que você aprende na escola é trabalho de muitas gerações. Receba essa herança, honre-a, acrescente a ela e, um dia, fielmente, deposite-a nas mãos de seus filhos.”*

Albert Einstein  
(1879-1955)

**Dedico esta dissertação à minha  
família, pelo apoio e encorajamento. A  
realização deste trabalho passa em  
grande parte por suas mãos.**

## **Agradecimentos**

Ao meu orientador, Fabio Gozzo, pela oportunidade de participar de seu grupo de pesquisa, pelas infindáveis e excelentes discussões sobre os mais diversos temas, pela disponibilidade e pelo prazer em partilhar seus conhecimentos. Também pela amizade e ótimo convívio, pela motivação e, sobretudo, pela incrível capacidade de transformar aquele “Fabio, acho que essa reação não deu muito certo” em “os resultados ficaram fantásticos!”.

Aos meus pais, Gisele e Vanderlei, pelo apoio incondicional, carinho e encorajamento nos momentos mais difíceis. Aos meus avós, Olga, Maria e Antônio, pelo amor e incentivo, por sempre acreditarem na minha capacidade. Ao meu avô Dirceu, meu primeiro “mentor”, por me inspirar desde pequeno a ter curiosidade e criatividade para solucionar os problemas (e consertar praticamente qualquer coisa). Aos meus tios Leslie e Wilson, Sandra e Edson, Heliana e Reinaldo, e aos primos mais próximos, pelo carinho e motivação. Em particular ao meu tio Alex, por me abrir as portas da M-8, nossa casa na moradia que nos rende piadas internas até hoje, e principalmente pela inspiração em procurar sempre fazer o que se gosta, o que para nós acabou significando seguir enfrentando os percalços da vida acadêmica.

À Caren, pelo amor, carinho, cuidado e dedicação, por me ouvir e consolar nos momentos difíceis e celebrar comigo a cada conquista. Por ouvir atentamente todas as minhas apresentações mesmo não entendendo uma palavra do que eu dizia e, principalmente, por sua habilidade quase sobrenatural de sempre me fazer sorrir.

Aos grandes amigos que tanto me apoiaram no decorrer dessa jornada, os limeirenses Afonso, André, Renan, Ricardo e Thiago, por entenderem aqueles milhares de programas que eu não pude comparecer porque tinha “um monte de coisas para fazer no lab” e ajustarem suas agendas para que ao menos às vezes pudéssemos colocar o papo em dia.

Um agradecimento ainda mais especial aos colegas de laboratório, aos ex-alunos Hugo e Mari; e atuais Tati, Renan, André e Renata, pelos aprendizados constantes e excelente convívio. Em particular, aos companheiros de salinha e de sorvete, Luana pelas risadas, discussões filosóficas e desabafos; e Állan, por tudo isso e mais as eternas discussões sobre orbitais, energias, modelagens, Linux, e orbitais novamente, afinal não foram poucos.

Ao Instituto de Química da UNICAMP, pela formação de excelência e infraestrutura. Aos funcionários da CPG, pela solicitude em sanar todas as dúvidas sobre cada uma das etapas que compuseram este trabalho.

Ao órgão de fomento CNPq, pela bolsa concedida.

## RESUMO

A técnica de ligação cruzada associada à espectrometria de massas (XL-MS) vem sendo utilizada com sucesso na resolução de estruturas de complexos proteicos. O método consiste na união covalente de dois ou mais resíduos reativos e espacialmente próximos da proteína através de um agente de ligação cruzada (ALC), que é uma molécula orgânica contendo dois ou mais grupos funcionais conectados por uma cadeia espaçadora. Os grupos funcionais são reativos frente às cadeias laterais de alguns resíduos da proteína, e se esse peptídeo modificado puder ser identificado, pode ser inferida uma restrição de distância entre os resíduos, indicando que, na estrutura terciária ou quaternária, a proximidade dos resíduos que reagiram é de no máximo o comprimento da cadeia espaçadora do ALC. Essa informação é valiosa para guiar uma modelagem computacional na obtenção de um modelo ou um conjunto de modelos que representem a estrutura da proteína em solução. Os ALC mais comuns utilizados em XL-MS são ésteres derivados de N-hidroxisuccinimida (NHS), em particular o DSS (suberato de dissuccinimidila), que é reativo frente a resíduos nucleofílicos, isto é, lisinas e, em menor escala, serinas. O uso de DSS possui uma série de limitações, que vão desde sua baixa solubilidade em solução aquosa até sua alta taxa de hidrólise e ao baixo rendimento da reação em baixas temperaturas. Para contornar essas limitações, foi desenvolvida uma nova química de ligação cruzada reativa também frente a resíduos nucleofílicos, utilizando o ácido adípico como ALC. Essa nova química, apresenta uma taxa de incorporação de moléculas de ácido adípico por proteína maior que a do DSS, mesmo contendo uma cadeia espaçadora menor. Além disso, a metodologia continua bastante reativa em baixas temperaturas, mantendo aproximadamente a mesma taxa de incorporação a 5, 15 e 25 °C, ao contrário da reatividade praticamente nula do DSS a 5 e 15 °C. Esses resultados mostram que a química desenvolvida traz diversas vantagens sobre a química de derivados de NHS, expandindo o uso da técnica de ligação cruzada para estruturas de proteínas e complexos em temperaturas muito mais baixas e, portanto, para proteínas que não são estáveis a temperatura ambiente.



## ABSTRACT

Chemical cross-linking coupled to mass spectrometry (XL-MS) has been successfully used to solve the structure of proteins complexes. The method consists in covalently linking two or more reactive and spatially close residues of the protein by a cross-linker, which is an organic molecule that has two or more functional groups connected by a spacer arm. The functional groups are able to react with the side chains of some of the protein's residues, and if the cross-linked peptide can be identified a distance can be inferred between the residues, indicating that in the tertiary (or quaternary) structure, both residues are spatially close within the reach of the spacer arm. This information is valuable to guide a computational modelling to obtain a model or a set of models that represents the structure of that protein in solution. The most common cross-linkers used in XL-MS are esters derived of N-hydroxysuccinimide, in particular DSS (disuccinimidyl suberate) which are reactive against nucleophilic residues, i.e. lysine and, in smaller scale, serine. The use of DSS has many limitations, which goes from its low solubility in aqueous medium to its tendency to hydrolysis and the reaction's low yield in lower temperatures. To circumvent these limitations, we developed a new cross-linking chemistry targeting nucleophilic residues, where adipic acid act as a cross-linker. This chemistry presents a higher incorporation rate than DSS, i.e. it incorporates more cross-linker molecules per protein than DSS, even having a smaller spacer arm. Moreover, this methodology is still very reactive at low temperatures, keeping approximately the same incorporation rate at 5, 15 and 25 °C, in contrast to the practically nonexistent reactivity of DSS reaction at 5 and 15 °C. These results show that this new cross-linking chemistry presents several advantages over NHS based cross-linkers, expanding its applications to structures of proteins and complexes in very low temperatures and, therefore, to proteins that are not stable at room temperatures.

## ÍNDICE DE FIGURAS

**Figura 1.** Ligação peptídica entre aminoácidos, na qual observa-se sua planaridade. As ligações N–C $\alpha$  e C $\alpha$ –C podem rotacionar, formando os ângulos  $\varphi$  e  $\psi$ , respectivamente. \_\_\_\_\_ 18

**Figura 2.** Representação do arranjo estrutural das estruturas secundárias. A) distribuição espacial dos átomos da cadeia principal e das cadeias laterais em uma  $\alpha$ -hélice e formação de ligações de hidrogênio entre os resíduos  $i$  e  $i+4$ , que estabilizam a conformação, além da representação somente da cadeia principal que compõe a hélice. B) distribuição dos átomos da cadeia principal e das cadeias laterais em uma folha- $\beta$ , mostrando estrutura em ziguezague, e formação das ligações de hidrogênio entre fitas paralelas que compõem a folha. \_\_\_\_\_ 19

**Figura 3.** Representação dos níveis estruturais de uma proteína, baseados na estrutura da hemoglobina (PDB 1C7D). A estrutura primária corresponde à sequência de resíduos de aminoácidos da cadeia proteica. Ligações de hidrogênio entre os grupos amida desses resíduos levam à formação de estruturas regulares do tipo  $\alpha$ -hélice e folha- $\beta$ . O arranjo tridimensional dos resíduos forma a estrutura terciária, mantida por interações entre as cadeias laterais dos mesmos. Interações entre estruturas terciárias podem levar à formação de estruturas supramoleculares, também chamadas de estruturas quaternárias. \_\_\_\_\_ 20

**Figura 4.** Estatísticas do número de estruturas de proteínas resolvidas e depositadas no PDB ao longo do tempo, já atualizada com dados parciais de 2017. A) número total de estruturas depositadas. B) Número de estruturas depositadas resolvidas pela técnica de difração de raios-X. C) Número de estruturas depositadas resolvidas pela técnica de ressonância magnética nuclear. D) Número de estruturas depositadas resolvidas pela técnica de microscopia eletrônica. \_\_\_\_\_ 22

**Figura 5.** Estatística do NYSGR (New York Structural Genomics Consortium) mostrando as taxas de sucesso em cada etapa da obtenção de um monocristal de uma proteína alvo. Em relação aos alvos clonados, a taxa de sucesso para a resolução da estrutura é de apenas 2% para a técnica de DRX. \_\_\_\_\_ 23

**Figura 6.** Estruturas supramoleculares depositadas no wwPDB. PRC2 (do inglês, Polycomb Repressive Complex 2) foi determinada por EM com 21 Å de resolução e combinado com XL-MS. A subunidade do ribossomo 39S foi determinada por EM com

4,1 Å de resolução e combinada com XL-MS. O complexo de RNA pol II foi resolvida a 16 Å por EM combinado com XL-MS. \_\_\_\_\_ 24

**Figura 7.** Representação esquemática dos métodos em proteômica estrutural obtidos através de técnicas baseadas em MS. \_\_\_\_\_ 26

**Figura 8.** Representação esquemática de um experimento típico de ligação cruzada com análise por MS para obtenção de dados estruturais de proteínas na forma de restrições espaciais de distância. 1) Reação da proteína alvo com o ALC desejado; 2) Digestão enzimática da proteína modificada, de forma a gerar peptídeos modificados e não-modificados; 3) Análise da mistura de peptídeos por LC-MS/MS, visando identificar os peptídeos contendo a modificação causada pela ligação cruzada, 4) Mapeamento das restrições de distância obtidas pelo experimento. 5) Uso do conjunto de restrições de distância como guia na modelagem computacional ou na seleção de modelos gerados que representem a estrutura da proteína alvo. \_\_\_\_\_ 28

**Figura 9.** Representação dos tipos de espécies formadas em experimentos de ligação cruzada. Em A) uma representação dessas espécies após a proteólise enzimática da proteína (ou complexo) alvo. Em B) uma representação dessas espécies ligadas na superfície da proteína antes da proteólise enzimática. \_\_\_\_\_ 29

**Figura 10.** Estrutura química de alguns ALCs. A e B) ALCs com sítios cliváveis; C e D) ALCs com grupos de afinidade; e E e F) ALCs heterobifuncionais fotoativáveis. 32

**Figura 11.** Série de agentes de ligação cruzada homólogos derivados de ésteres de NHS com diferentes tamanhos de cadeia alifática espaçadora. O ALC mais utilizado, DSS, tem  $n=6$  e uma cadeia espaçadora que mede 11,4 Å. \_\_\_\_\_ 34

**Figura 12.** Representação esquemática da realização do experimento de ligação cruzada utilizando o ácido adípico como ALC. Na parte de baixo da figura, estão representadas as estruturas das moléculas utilizadas nessas reações. \_\_\_\_\_ 38

**Figura 13.** Representação do mecanismo de formação de uma espécie de ligação cruzada utilizando o DSS como ALC sendo atacado por um grupo nucleofílico de um resíduo de lisina. \_\_\_\_\_ 43

**Figura 14.** Representação da etapa de ativação dos grupos ácidos nas duas metodologias desenvolvidas no grupo de pesquisa. A) ativação dos grupos ácidos dos resíduos de ácido aspártico de glutâmico e uso da 1,6-hexanodiamina como ALC, formando espécies de ligação cruzada entre esses resíduos eletrofílicos; B) ativação dos grupos ácidos de um diácido usado como ALC e a reação da espécie ativada com

resíduos de lisina, formando espécies de ligação cruzada entre esses resíduos nucleofílicos. \_\_\_\_\_ 45

**Figura 15.** Representação do mecanismo de formação de produtos de ligação cruzada utilizando ácido adípico como ALC. A) mecanismo da dupla ativação do ácido adípico com EDC e HOBt; B) mecanismo da reação da espécie ativada com resíduos de lisina para formação das espécies de ligação cruzada. \_\_\_\_\_ 46

**Figura 16.** Espectros deconvoluídos de ESI(+)-MS de mioglobina intacta na presença e ausência dos ativadores do ácido adípico. Nota-se que a dupla ativação do ácido adípico é realmente necessária para sua incorporação na proteína. \_\_\_\_\_ 48

**Figura 17.** Ampliação do espectro deconvoluído de ESI(+)-MS da reação de mioglobina com ácido adípico duplamente ativado. Para cada incorporação de ALC, observa-se sinais correspondentes a diferenças de massa de 18 Da, indicando a formação de espécies de ligação cruzada do tipo zero-length. No quadro interno, uma ampliação ainda maior mostrando a massa correspondente desses sinais. \_\_\_\_\_ 49

**Figura 18.** Mecanismo de reação para formação de produtos de ligação cruzada do tipo zero-length. A) ativação dos resíduos ácidos (representado pelo ácido glutâmico) com EDC e HOBt; B) reação do resíduo ácido ativado com resíduo nucleofílico próximo (representado por lisina) para formação do produto de ligação cruzada. \_ 50

**Figura 19.** Gráfico apresentando a abundância relativa dos resíduos de aminoácidos no conjunto de aproximadamente 60 milhões de proteínas do banco de dados TrEMBL. \_\_\_\_\_ 51

**Figura 20.** Espectros deconvoluídos de ESI(+)-MS dos produtos de ligação cruzada com várias adições de EDC a cada 30 minutos. Observa-se que não há aumento do rendimento de formação de ligação cruzada com adições extras de EDC ao meio reacional. \_\_\_\_\_ 53

**Figura 21.** Espectros deconvoluídos de ESI(+)-MS da reação da mioglobina com diferentes excessos de ácido adípico. Observa-se que a taxa de incorporação média de moléculas de ALC aumenta com os excessos molares de ácido adípico, atingindo valores ideais com 100 e 300 vezes de excesso. Em todas as reações as proporções de EDC e HOBt foram mantidas em 2:1 e 8:1 em relação ao ALC, respectivamente. \_\_\_\_\_ 55

**Figura 22.** Espectros deconvoluídos de ESI(+)-MS da reação da mioglobina utilizando diferentes períodos de ativação do ácido adípico. Observa-se que o padrão de

incorporação do ALC não se altera significativamente em nenhuma das condições, apresentando uma incorporação em média com ativação de 60 minutos. \_\_\_\_\_ 58

**Figura 23.** Espectros deconvoluídos de ESI(+)-MS de mioglobina comparando as reações de ligação cruzada utilizando o DSS e o ácido adípico como ALC. Nota-se que, mesmo na temperatura ótima da reação com DSS, a taxa de incorporação do ALC por molécula da proteína é consideravelmente menor que a do ácido adípico. 60

**Figura 24.** Espectros deconvoluídos de ESI(+)-MS de mioglobina nas reações de ligação cruzada para avaliação da influência da temperatura na taxa de incorporação de DSS. Observa-se que, com o decréscimo da temperatura, a incorporação do ALC cai drasticamente. \_\_\_\_\_ 62

**Figura 25.** Espectros deconvoluídos de ESI(+)-MS de mioglobina nas reações de ligação cruzada para avaliação da influência da temperatura na taxa de incorporação do ácido adípico. Observa-se que, mesmo com o decréscimo da temperatura, a taxa de incorporação do ALC se mantém praticamente inalterada. \_\_\_\_\_ 63

**Figura 26.** Exmplos de espectros de íons fragmentos anotados para espécies de ligação cruzada. A) espécies XL entre lisinas através de reação com DSS; B) espécies XL entre lisinas através de reação com ácido adípico; C) espécies ZL entre lisina e ácido glutâmico através de reação com ácido adípico. \_\_\_\_\_ 65

**Figura 27.** Mapa das ligações cruzadas identificadas para a mioglobina à 25 e 5 °C com DSS (A) e ácido adípico (B). Nota-se a formação de mais espécies nas duas reações em que foi utilizado o ácido adípico como ALC, mesmo em temperatura reduzida. Também é possível notar uma cobertura maior da sequência da proteína. \_\_\_\_\_ 66

**Figura 28.** Mapa das ligações cruzadas identificadas para a mioglobina à 25 e 5 °C com ácido adípico como ALC em pH=7 (A) e pH=6,5 (B). Quando comparadas as mesmas temperaturas, observa-se a formação de um maior número de espécies com pH=6,5. Já quando duas reações no mesmo pH são comparadas, são observadas menos espécies nas reações em temperaturas mais baixas. \_\_\_\_\_ 68

**Figura 29.** Espectros deconvoluídos de ESI(+)-MS de citocromo c comparando as reações de ligação cruzada utilizando o DSS e o ácido adípico como ALC. Mesmo na temperatura ótima da reação com DSS, sua taxa de incorporação por molécula da proteína é consideravelmente menor que a do ácido adípico. \_\_\_\_\_ 71

**Figura 30.** Espectros deconvoluídos de ESI(+)-MS de citocromo c nas reações de ligação cruzada para avaliação da influência da temperatura na taxa de incorporação

de DSS. Observa-se que, com o decréscimo da temperatura, o rendimento da reação é muito reduzido. \_\_\_\_\_ 72

**Figura 31.** Espectros deconvoluídos de ESI(+)-MS de citocromo c nas reações de ligação cruzada para avaliação da influência da temperatura na taxa de incorporação de ácido adípico. Nota-se que, com o decréscimo da temperatura, a incorporação do ALC se mantém praticamente constante. A 5 °C, há a incorporação também de 10 espécies do ALC, mas apenas do tipo dead-end, sendo que com 7 incorporações, em média, há também a formação de espécies de ligação cruzada. \_\_\_\_\_ 73

## ÍNDICE DE TABELAS

**Tabela 1.** Número de restrições totais e divididas por espécie obtidas nos experimentos de ligação cruzada com ácido adípico em diferentes temperaturas e pH. É mostrado também a composição do número total da espécie de acordo com os resíduos reativos. Por exemplo, em pH 6,5 / 5 °C há 42 restrições, 18 do tipo XL e 24 do tipo ZL. Dentro das 18 espécies XL, 13 XL são entre lisinas (K-K), 5 entre lisina e serina (K-S) e nenhuma entre serinas (S-S). \_\_\_\_\_ 69

## LISTA DE ABREVIações E SIGLAS

<b>ALC</b>	Agente de Ligação Cruzada
<b>CD</b>	Dicroísmo Circular
<b>Cryo-EM</b>	Crio-Microscopia Eletrônica
<b>Da</b>	Dalton (1 Da = $1,661 \times 10^{-24}$ g)
<b>DMF</b>	N,N-dimetilformamida
<b>DRX</b>	Difração de Raios-X
<b>DSS</b>	Suberato de N,N-dissuccinimidila
<b>EM</b>	Microscopia Eletrônica
<b>ESI</b>	Ionização por <i>Eletrospray</i>
<b>FRET</b>	Transferência de Energia Ressonante por Fluorescência
<b>FTICR</b>	Ressonância Ciclotrônica de Íons com Transformada de Fourier
<b>HDX</b>	Troca Hidrogênio/Deutério
<b>IM</b>	Mobilidade Iônica
<b>LIT</b>	Analizador de massas do tipo ion trap linear
<b>MALDI</b>	Ionização por Dessorção a Laser Auxiliada por Matriz
<b>MS</b>	Espectrometria de Massas
<b>MS<sup>n</sup></b>	Espectrometria de Massas Sequencial
<b>NHS</b>	N-hidroxisuccinimida
<b>NYSGR</b>	<i>New York Structural Genomics Research Consortium</i>
<b>PDB</b>	<i>Protein Data Bank</i>
<b>Q</b>	Analizador de massas do tipo quadrupolo
<b>RMN</b>	Ressonância Magnética Nuclear
<b>SAXS</b>	Espalhamento de Raios-X a Baixos Ângulos
<b>SIM-XL</b>	<i>Spectrum Identification Machine for Cross-Linked peptides</i>
<b>TOF</b>	Analizador de massas do tipo tempo de voo
<b>UPLC</b>	Cromatografia Líquida de Ultra Alta Eficiência
<b>UV</b>	Ultravioleta
<b>wwPDB</b>	<i>World Wide Protein Data Bank</i>
<b>XL</b>	Produtos de Ligação Cruzada do tipo inter ou intramoleculares
<b>XL-MS</b>	Ligação Cruzada associada à Espectrometria de Massas
<b>ZL</b>	Produtos de Ligação Cruzada do tipo <i>zero-length</i>



## SUMÁRIO

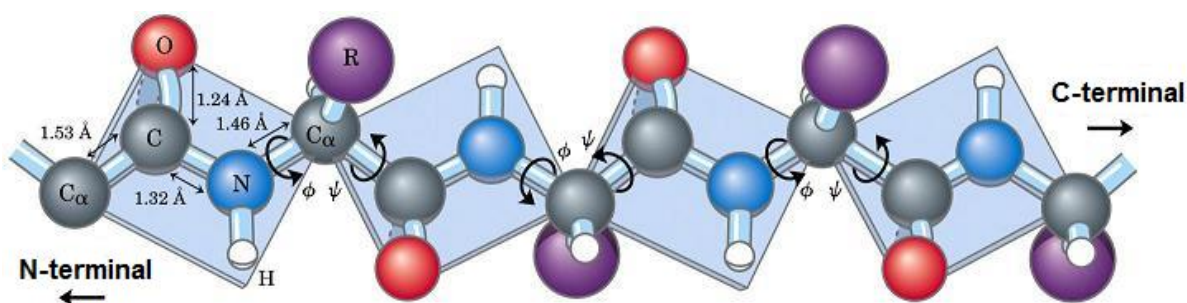
1. INTRODUÇÃO	18
1.1. Estrutura de Proteínas	18
1.2. Métodos para caracterização estrutural de proteínas	21
1.3. Espectrometria de massas no estudo estrutural de proteínas	24
1.4. Ligação cruzada associada à espectrometria de massas no estudo estrutural de proteínas	26
1.5. Desafios e avanços na área de ligação cruzada	29
1.6. Uso de derivados de ésteres de NHS como ALCs e suas limitações	33
2. OBJETIVOS	36
2.1. Objetivo Geral	36
2.2. Objetivos Específicos	36
3. PROCEDIMENTO EXPERIMENTAL	37
3.1. Ensaios de ligação cruzada utilizando o ácido adípico como ALC	37
3.2. Ensaios de ligação cruzada utilizando o DSS como ALC	38
3.3. Análise de proteína intacta modificada pelo ALC por MS	39
3.4. Proteólise enzimática	39
3.5. Análise dos peptídeos modificados pelo ALC por MS	40
3.6. Tratamento dos dados e identificação das espécies de ligação cruzada	40
4. RESULTADOS E DISCUSSÃO	42
4.1. Uso do DSS como ALC reativo frente a resíduos nucleofílicos e possíveis alternativas	42
4.2. Dupla ativação de grupos ácidos e uso do ácido adípico como ALC	44
4.3. Otimização da metodologia de ligação cruzada utilizando o ácido adípico	52
4.4. Estudo comparativo entre os experimentos de ligação cruzada utilizando o ácido adípico frente o uso de DSS	59
4.5. Avaliação da metodologia proposta através da aplicação à proteína citocromo c	70
5. CONCLUSÕES	75
6. REFERÊNCIAS	76

## 1. INTRODUÇÃO

### 1.1. Estrutura de Proteínas

As proteínas constituem o grupo de macromoléculas mais diverso, complexo e versátil entre as biomoléculas encontradas na natureza, e são formadas pela união covalente entre aminoácidos através de ligações peptídicas (amidas).

Essa ligação, formada pela reação do grupo amina de um aminoácido com o grupo carboxila de outro, apresenta caráter parcial de dupla ligação devido à ressonância do par de elétrons do orbital p do nitrogênio com o par de elétrons  $\pi$  da carbonila, sendo, portanto, planar<sup>1</sup>. Devido à hibridização  $sp^2$  do carbono da carbonila, o carbono vizinho a ela – chamado de  $C\alpha$  – se encontra também no mesmo plano. No entanto, como as ligações  $N-C\alpha$  e  $C\alpha-C$  são simples, as mesmas possuem liberdade de rotação, e os ângulos adotados por elas em relação ao plano são denominados ângulos  $\phi$  ( $\phi$ ) e  $\psi$  ( $\psi$ ), respectivamente (Figura 1).

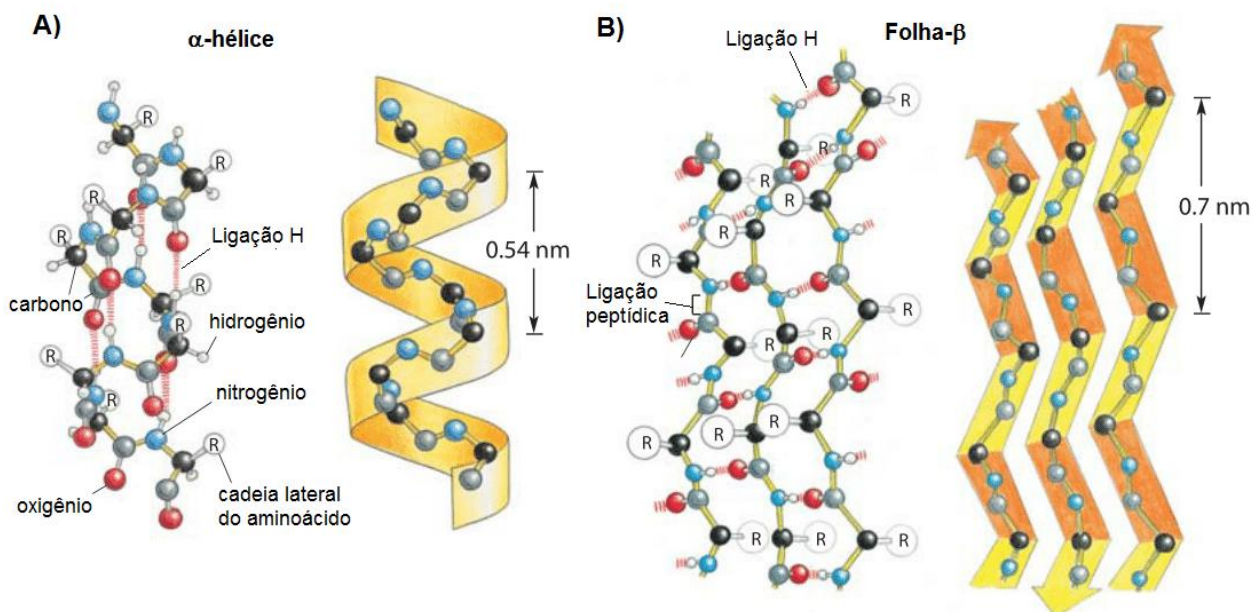


**Figura 1.** Ligação peptídica entre aminoácidos, na qual observa-se sua planaridade. As ligações  $N-C\alpha$  e  $C\alpha-C$  podem rotacionar, formando os ângulos  $\phi$  e  $\psi$ , respectivamente.

A estrutura de uma proteína pode ser sistematicamente dividida em quatro níveis organizacionais, sendo que o primeiro nível, também chamado de estrutura primária, corresponde à sequência de aminoácidos que a compõem. Essa sequência é única para cada proteína e define tanto sua estrutura tridimensional quanto sua função. A formação de ligações de hidrogênio entre os grupos amida dos resíduos leva a estruturas regulares e definidas do tipo  $\alpha$ -hélice ou folha- $\beta$ , sendo essas chamadas de estruturas secundárias.

As  $\alpha$ -hélices apresentam estrutura helicoidal na qual cada resíduo corresponde a uma curva de  $100^\circ$ , ou seja, uma volta completa da hélice possui 3,6 resíduos. O oxigênio da carbonila de cada um deles faz uma ligação de hidrogênio

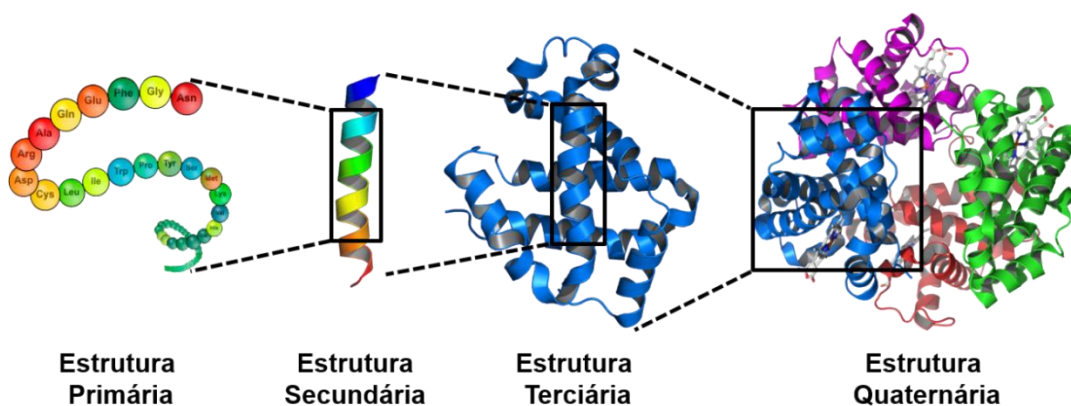
com o hidrogênio do grupo N–H de outro resíduo localizado 4 aminoácidos a frente ( $i \rightarrow i+4$ ). As folhas- $\beta$  apresentam estrutura ziguezague, na qual as ligações de hidrogênio são formadas entre segmentos adjacentes da folha ou, mais precisamente, entre as fitas que compõem a folha (Figura 2)<sup>1,2</sup>.



**Figura 2.** Representação do arranjo estrutural das estruturas secundárias. A) Distribuição espacial dos átomos da cadeia principal e das cadeias laterais em uma  $\alpha$ -hélice e formação de ligações de hidrogênio entre os resíduos  $i$  e  $i+4$ , que estabilizam a conformação, além da representação somente da cadeia principal que compõe a hélice. B) Distribuição dos átomos da cadeia principal e das cadeias laterais em uma folha- $\beta$ , mostrando estrutura em ziguezague, e formação das ligações de hidrogênio entre fitas paralelas que compõem a folha.

Algumas regiões das proteínas apresentam estruturas que não são regulares como as  $\alpha$ -hélices e folhas- $\beta$ , denominadas *turns* e *loops*. Os *turns* são pequenos segmentos de aminoácidos que ligam duas estruturas secundárias. Já os *loops* são segmentos com um maior número de resíduos que, por não possuírem arranjo regular e fortes interações intermoleculares, formam regiões geralmente mais flexíveis na estrutura da proteína.<sup>2</sup> Essas regiões desordenadas podem ser responsáveis por diversos papéis biológicos das proteínas nos organismos, como a interação com ligantes que podem causar mudanças conformacionais ativando ou desativando uma enzima, ou ainda com os próprios substratos quando a proteína está efetuando sua função<sup>3</sup>.

O arranjo tridimensional de todos os átomos da sequência da proteína, já contendo o arranjo entre as estruturas secundárias, é chamado de estrutura terciária. Nesse nível estrutural há a presença de interações entre as cadeias laterais dos resíduos de aminoácidos de natureza diversa, desde ligações de hidrogênio a pontes de dissulfeto, interações eletrostáticas e outras interações de Van der Waals. A interação entre proteínas leva à formação de estruturas supramoleculares, também chamadas de complexos proteicos, que correspondem ao nível quaternário de organização estrutural<sup>1,2</sup>. Uma representação dos quatro níveis estruturais é mostrada na Figura 3.



**Figura 3.** Representação dos níveis estruturais de uma proteína, baseados na estrutura da hemoglobina (PDB 1C7D). A estrutura primária corresponde à sequência de resíduos de aminoácidos da cadeia proteica. Ligações de hidrogênio entre os grupos amida desses resíduos levam à formação de estruturas regulares do tipo  $\alpha$ -hélice e folha- $\beta$ . O arranjo tridimensional dos resíduos forma a estrutura terciária, mantida por interações entre as cadeias laterais dos mesmos. Interações entre estruturas terciárias podem levar à formação de estruturas supramoleculares, também chamadas de estruturas quaternárias.

A enorme diversidade encontrada entre proteínas se deve ao número praticamente infinito de combinações entre as sequências de aminoácidos, bem como a ampla faixa de tamanhos e formas que apresentam, podendo ir de poucos kDa a alguns MDa<sup>2</sup>. Essa diversidade é responsável também pela versatilidade dessa classe de moléculas, já que as proteínas apresentam uma extensa gama de funções que vão desde a catálise de reações até o transporte de substâncias e defesa do organismo, por exemplo<sup>1</sup>.

A função biológica de uma proteína está diretamente ligada à sua estrutura e dinâmica, sendo que em muitos casos é a interação com ligantes e/ou outras

proteínas que induzem essa estrutura para uma conformação ativa, necessária para o desempenho de sua atividade<sup>4</sup>. Desse modo, essas interações são vitais para o funcionamento e manutenção de qualquer processo celular e, conseqüentemente, para a vida de qualquer organismo<sup>5</sup>, ilustrando a importância de estudos acerca dessas interações e das estruturas das proteínas como um todo.

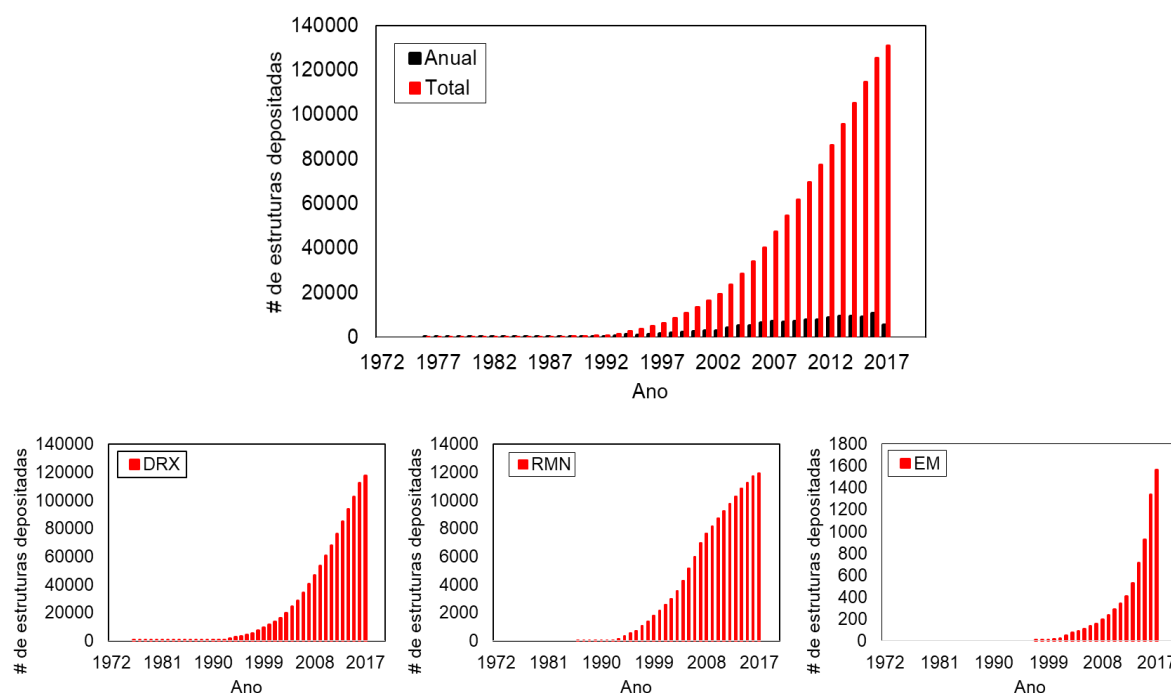
## **1.2. Métodos para caracterização estrutural de proteínas**

As técnicas para determinação estrutural de proteínas podem ser divididas em três grupos, dependendo da resolução que oferecem. Técnicas de baixa resolução normalmente fornecem informações mais restritas, principalmente a respeito de partes da proteína, como a média de estruturas secundárias que a compõem, a superfície exposta ao solvente ou regiões de interação entre proteínas, e englobam as espectroscopias, como Dicroísmo Circular (CD), Fluorescência no UV, Transferência de Energia Ressonante por Fluorescência (FRET), entre outras. Essas técnicas tendem a ser de mais fácil acesso, além de relativamente simples e rápidas.<sup>6,7</sup>

Técnicas de média resolução normalmente fornecem informações a respeito da estequiometria, além da posição e orientação dos componentes do sistema, e englobam técnicas como o espalhamento de raios-X a baixos ângulos (SAXS) e a microscopia eletrônica (EM)<sup>6,7</sup>. No entanto, avanços recentes em criomicroscopia eletrônica (Cryo-EM) para partículas únicas, detectores e processamento de imagens levaram ao desenvolvimento acelerado da técnica, que atualmente já é capaz de resolver estruturas de proteínas em alta resolução<sup>8,9</sup>.

Técnicas de resolução atômica, ou alta resolução, fornecem informações a respeito das posições e orientações dos átomos do sistema. Essas técnicas englobam a Difração de raios-X (DRX), a Ressonância magnética nuclear (RMN) e, mais recentemente e já citada, a Cryo-EM. Em geral são de acesso mais restrito devido à complexidade da instrumentação utilizada. A técnica de DRX foi utilizada pela primeira vez na caracterização estrutural de proteínas na década de 1950, na resolução das estruturas da mioglobina e da hemoglobina<sup>10,11</sup>, se tornando desde então a principal técnica para esse tipo de estudo<sup>12</sup>. No caso de RMN, a primeira determinação estrutural de uma proteína ocorreu na década de 1980, e desenvolvimentos na instrumentação contribuíram para a difusão no uso da técnica para caracterização de proteínas<sup>13</sup>. Atualmente, essas duas técnicas de alta resolução respondem por 98% de todas as estruturas de proteínas resolvidas e depositadas no PDB (*Protein Data*

Bank)<sup>14</sup>, sendo que 88% foram obtidas por DRX e 10% por RMN. Os outros 2% das estruturas são advindas das técnicas de EM (Figura 4).

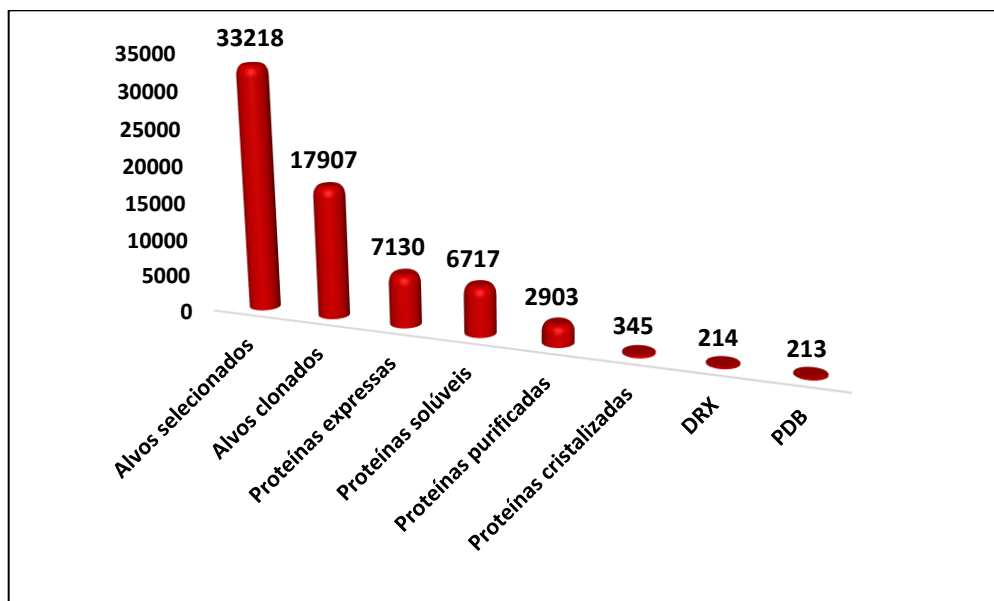


**Figura 4.** Estatísticas do número de estruturas de proteínas resolvidas e depositadas no PDB ao longo do tempo, já atualizada com dados parciais de 2017. A) número total de estruturas depositadas. B) Número de estruturas depositadas resolvidas pela técnica de difração de raios-X. C) Número de estruturas depositadas resolvidas pela técnica de ressonância magnética nuclear. D) Número de estruturas depositadas resolvidas pela técnica de microscopia eletrônica<sup>14</sup>.

Apesar das técnicas de alta resolução serem amplamente utilizadas para a caracterização estrutural, há uma série de limitações em seu uso, o que faz com que a maioria das proteínas não sejam passíveis de serem estudadas por elas. No caso de Cryo-EM, há uma limitação a respeito do tamanho da proteína a ser estudada, atualmente na ordem de centenas de kDa. Para RMN, essa limitação é ainda mais restritiva, sendo que proteínas acima de 40 kDa já fornecem espectros extremamente complexos e que inviabilizam sua resolução. Além disso, as proteínas analisadas por RMN devem ser estáveis em solução de tampões que não interfiram nas análises por longos períodos, que podem chegar a vários dias<sup>6</sup>.

No caso de DRX, o grande gargalo da técnica reside na dificuldade de obtenção de monocristais. De acordo com uma estatística do *New York Structural Genomics Research Consortium* (NYSGR), a grande maioria dos alvos selecionados

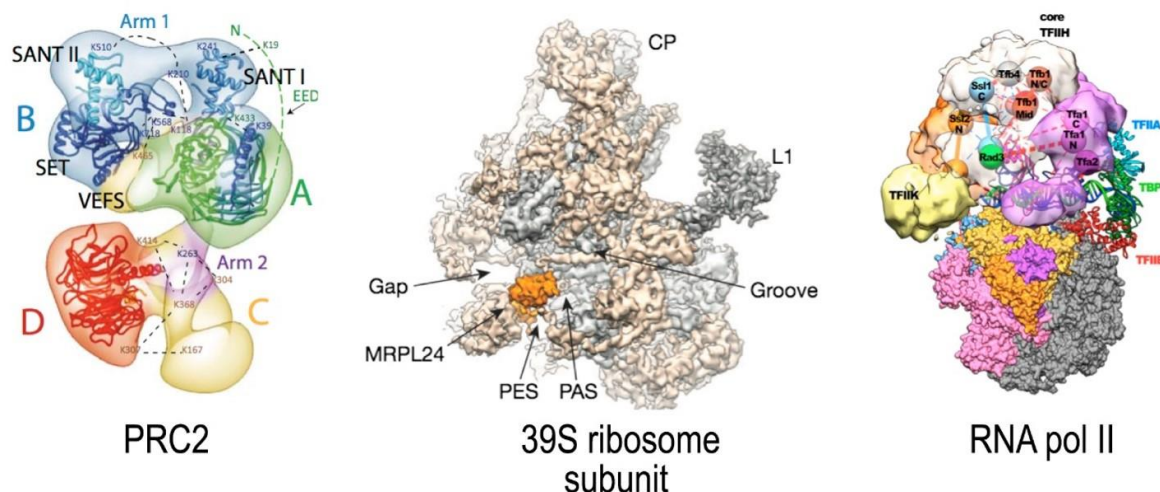
não chega às etapas de difração, sendo que a taxa de sucesso em relação ao número de proteínas clonadas é próxima de apenas 2% (Figura 5).



**Figura 5.** Estatística do NYSGR (New York Structural Genomics Consortium) mostrando as taxas de sucesso em cada etapa da obtenção de um monocristal de uma proteína alvo. Em relação aos alvos clonados, a taxa de sucesso para a resolução da estrutura é de apenas 2% para a técnica de DRX.

Como consequência dessas limitações apresentadas pelas técnicas de alta resolução disponíveis atualmente, a utilização de abordagens integrativas na obtenção de informações acerca da estrutura das proteínas constitui uma alternativa importante para estudos estruturais. Nesse tipo de abordagem são empregadas diversas técnicas como troca hidrogênio/deutério (HDX), SAXS, FRET, ligação cruzada associada à espectrometria de massas (XL-MS), dentre outras, em que cada técnica fornece um conjunto de informações a respeito da proteína de interesse que, quando unidas e combinadas com métodos computacionais, podem ser usadas na obtenção de modelos moleculares que representem os dados obtidos<sup>15,16</sup>. A Figura 6 mostra algumas estruturas supramoleculares obtidas através de abordagens integrativas e depositadas no wwPDB (*world wide protein data bank*)<sup>17</sup>, outro importante banco de dados de proteínas que desde o início da década de 2010 passou a aceitar depósitos de estruturas advindas dessas abordagens.





**Figura 6.** Estruturas supramoleculares depositadas no wwPDB<sup>17</sup>. PRC2 (do inglês, *Polycomb Repressive Complex 2*) foi determinada por EM com 21 Å de resolução e combinado com XL-MS. A subunidade do ribossomo 39S foi determinada por EM com 4,1 Å de resolução e combinada com XL-MS. O complexo de RNA pol II foi resolvida a 16 Å por EM combinado com XL-MS.

Dado a necessidade de diferentes metodologias para o estudo de uma gama mais abrangente de proteínas, diferentes técnicas vem sendo desenvolvidas ao longo das últimas décadas sendo uma das mais atrativas a espectrometria de massas (MS).

### 1.3. Espectrometria de massas no estudo estrutural de proteínas

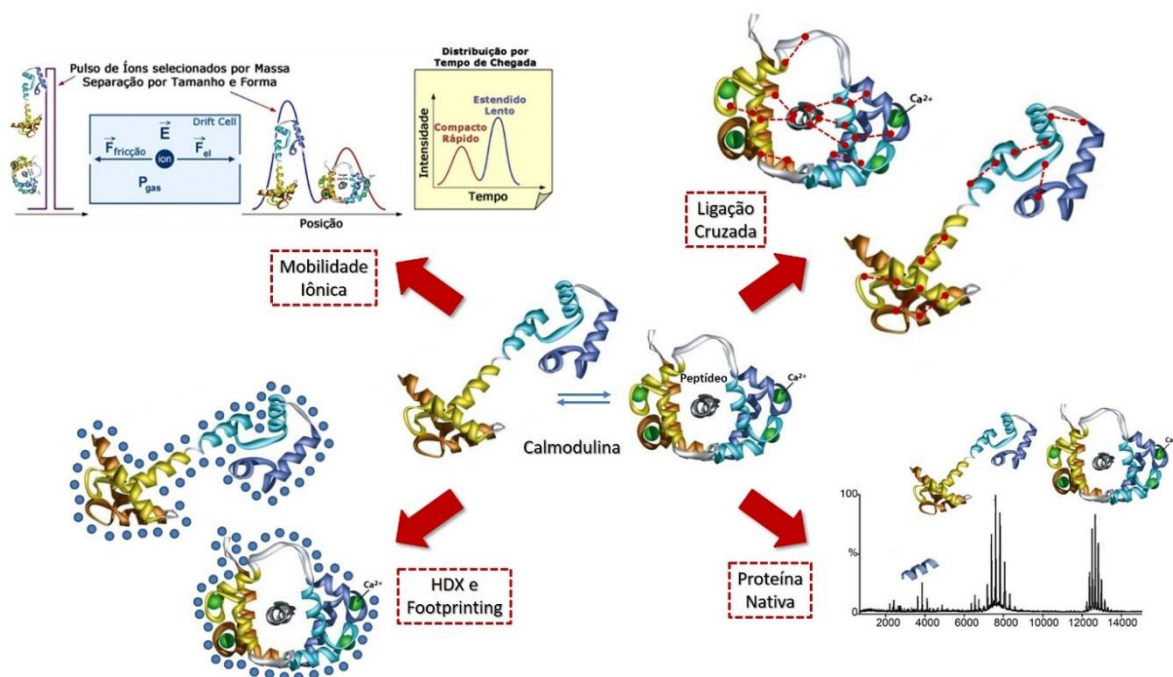
A espectrometria de massas consiste no estudo de íons em fase gasosa, sendo que uma de suas principais aplicações consiste na caracterização estrutural de compostos. Sua utilização no estudo de macromoléculas, no entanto, foi bastante limitada até o final da década de 1980, já que as fontes de ionização disponíveis até então eram inviáveis para essa aplicação<sup>18</sup>. Com o advento das técnicas de ionização suaves ESI (*Electrospray Ionization*)<sup>19</sup> e MALDI (*Matrix-Assisted Laser Desorption Ionization*)<sup>20</sup>, se tornou possível a ionização de biomoléculas, permitindo a análise de peptídeos, proteínas e ácidos nucleicos sem derivação ou fragmentação<sup>19</sup>.

O desenvolvimento dessas técnicas impulsionou uma grande e rápida evolução dos analisadores de massas associados, levando à disponibilização comercial de instrumentos com capacidade de experimentos de MS sequenciais (MS<sup>n</sup>), possuindo plataformas de diversas geometrias combinando diferentes analisadores, como quadrupolo – tempo de voo (Q-TOF), tempo de voo – tempo de



voo (TOF-TOF), ion trap linear (LIT), ressonância ciclotrônica de íons com transformada de Fourier (FTICR)<sup>21</sup>, ion trap – orbitrap (LTQ-Orbitrap)<sup>22</sup>, quadrupolo – orbitrap (Q-Orbitrap)<sup>23</sup>, entre outros. Desde então, com instrumentos de desempenho cada vez melhor, o uso da espectrometria de massas na análise de proteínas evoluiu extensivamente, a ponto de se tornar rotineiro no sequenciamento e identificação de proteínas, determinação da massa molecular de proteínas e complexos proteicos intactos, determinação de parceiros de interação, identificação e localização de modificações pós-traducionais e quantificação absoluta e relativa de proteínas<sup>24–30</sup>. Somando-se a isso, o desenvolvimento de metodologias envolvendo cromatografia líquida de ultra alta eficiência (UPLC) acoplada a MS para análise de misturas complexas de peptídeos resultantes de digestão enzimática de uma grande variedade de proteínas, além do crescimento das ferramentas de bioinformática para buscas em bancos de dados, possibilitou a realização de estudos de proteômica em larga escala, chamados também de proteômica *shotgun*.

A técnica de MS possui algumas vantagens que lhe são intrínsecas, colaborando para o grande aumento de seu uso na caracterização de proteínas, como alta sensibilidade (de modo que não é necessária grande quantidade das proteínas-alvo), rapidez, versatilidade, facilidade de operação e aplicabilidade universal, visto que não há uma restrição conhecida de tamanho ou qualquer outra característica da proteína a ser estudada. Além disso, a técnica também dispensa a necessidade de elevados graus de pureza da amostra, condição necessária nas técnicas de alta resolução como RMN, DRX e Cryo-EM. Essas características únicas acabaram impulsionando o desenvolvimento de novas técnicas e metodologias baseadas em MS para a análise de estruturas superiores de proteínas (Figura 7), como HDX<sup>31</sup> e mapeamento radicalar oxidativo (*Footprinting*)<sup>32</sup>, Mobilidade Iônica (IM-MS)<sup>33</sup> e, mais recentemente, ligação cruzada associada à espectrometria de massas (XL-MS)<sup>34</sup>.



**Figura 7.** Representação esquemática dos métodos em proteômica estrutural obtidos através de técnicas baseadas em MS.

#### 1.4. Ligação cruzada associada à espectrometria de massas no estudo estrutural de proteínas

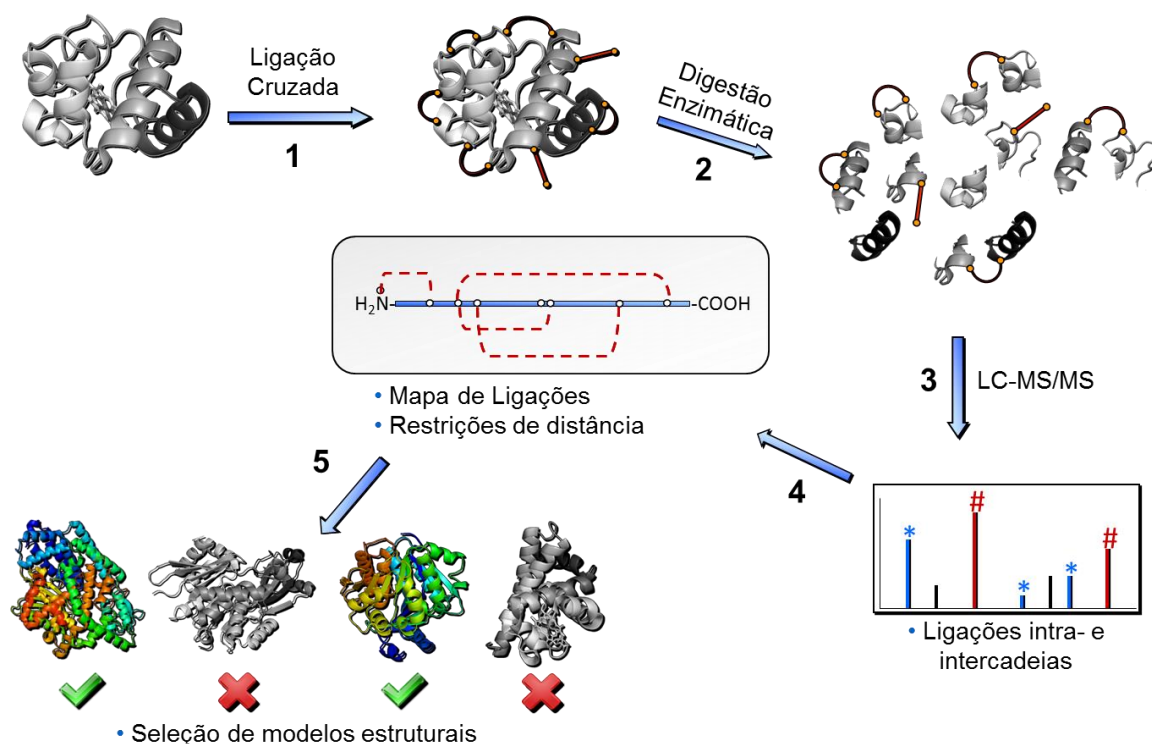
O fenômeno de ligação cruzada compreende a união covalente entre duas espécies. Essa união pode ser independente do uso de outros compostos, como no caso das ligações de dissulfeto entre resíduos de cisteína, ou pode ser realizada por intermédio de um agente de ligação cruzada (ALC), sendo que as espécies envolvidas nos experimentos de ligação cruzada podem pertencer a diferentes classes, como peptídeos, proteínas, moléculas orgânicas, ácidos nucleicos e até mesmo partículas sólidas<sup>35</sup>. A técnica era utilizada inicialmente para a estabilização de proteínas, imobilização de macromoléculas em suportes sólidos, acoplamento de duas ou mais moléculas de proteína e extensão de interações proteína-proteína, e só começou a ser empregada associada à espectrometria de massas para a obtenção de informações estruturais acerca das proteínas nos anos 2000<sup>35–37</sup>. Nos últimos anos, essa metodologia tem sido aplicada na determinação de regiões de interações entre proteínas, como é o caso de complexos proteicos<sup>38–40</sup> e interação entre proteínas e ligantes<sup>41–43</sup>.

Na aplicação de ligação cruzada em estruturas superiores de proteínas, as cadeias laterais de resíduos de aminoácidos espacialmente próximos reagem com as extremidades reativas dos ALCs, que funcionam como réguas moleculares

forneendo informações a respeito da distância em que os resíduos reativos se encontram na estrutura. Os ALCs são compostos orgânicos multifuncionais, contendo em geral dois grupos reativos unidos por uma cadeia espaçadora de comprimento variável, e são esses grupos reativos que conferem a cada ALC uma determinada especificidade em relação a alguns resíduos que constituem as proteínas<sup>44</sup>. Dependendo se os grupos reativos são iguais ou diferentes, os ALCs podem ser classificados como homobifuncionais e heterobifuncionais, respectivamente, o que permite maior versatilidade na especificidade de cada reagente.

Uma vez que a ligação cruzada tenha ocorrido e o ALC tenha se ligado à cadeia lateral dos resíduos reativos, cria-se uma modificação de massa nessa proteína ou peptídeo, e é na identificação dessas espécies que a análise por MS é empregada. A análise dos produtos de ligação cruzada permite obter informações acerca da distância entre os resíduos reativos. Mais precisamente, são obtidas restrições de distância entre esses resíduos indicando que, na estrutura nativa, eles se encontram separados por no máximo o comprimento da cadeia espaçadora do ALC utilizado.

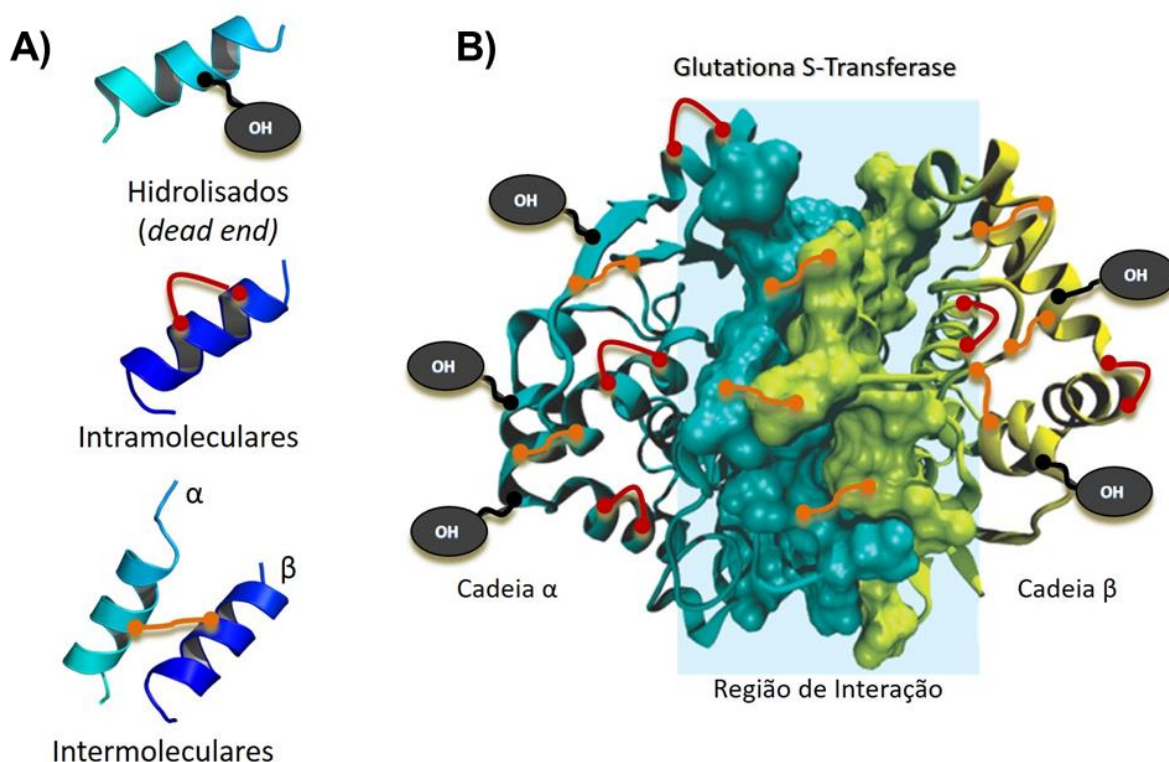
Essa análise é feita submetendo-se a proteína ou complexo proteico modificados pelos ALC a proteólise enzimática e identificando-se os peptídeos modificados por LC-MS e LC-MS/MS, sendo que a mistura de peptídeos passa por uma separação cromatográfica seguida de análise por MS de forma contínua. Os íons de peptídeos são automaticamente submetidos a experimentos de MS/MS durante a corrida cromatográfica, permitindo identificá-los e localizar as modificações causadas pelo ALC (Figura 8). Essa abordagem é semelhante aos métodos de proteômica *shotgun* empregados em análises de misturas complexas de proteínas.



**Figura 8.** Representação esquemática de um experimento típico de ligação cruzada com análise por MS para obtenção de dados estruturais de proteínas na forma de restrições espaciais de distância. 1) Reação da proteína alvo com o ALC desejado; 2) Digestão enzimática da proteína modificada, de forma a gerar peptídeos modificados e não-modificados; 3) Análise da mistura de peptídeos por LC-MS/MS, visando identificar os peptídeos contendo a modificação causada pela ligação cruzada, 4) Mapeamento das restrições de distância obtidas pelo experimento. 5) Uso do conjunto de restrições de distância como guia na modelagem computacional ou na seleção de modelos gerados que representem a estrutura da proteína alvo.

Os produtos da proteólise enzimática modificados com o ALC podem ser classificados em três tipos (Figura 9): 1) espécies intermoleculares, nas quais as duas extremidades do ALC reagiram com dois resíduos que, após a digestão enzimática, estão contidos em diferentes peptídeos; 2) espécies intramoleculares, nas quais as duas extremidades do ALC também reagiram com dois resíduos, mas que se encontram no mesmo peptídeo após a digestão enzimática; e 3) espécies parcialmente hidrolisadas ou de *dead-end*, nas quais apenas uma das extremidades do ALC reagiu com a cadeia lateral de um resíduo da proteína, enquanto que a outra extremidade sofreu hidrólise no meio, seja por labilidade do grupo reativo ou simplesmente pela proteína não possuir outro resíduo reativo dentro do alcance do ALC. Cada uma dessas espécies fornece um conjunto diferente de informações

acerca da estrutura da proteína ou complexo proteico. As espécies de *dead-end* trazem informações sobre a acessibilidade ao solvente do resíduo reativo; as espécies intramoleculares mostram que os resíduos reativos contidos no mesmo peptídeo (portanto, próximos na estrutura primária) apresentam também proximidade na estrutura terciária ou quaternária; já as espécies intermoleculares revelam que, embora distantes na estrutura primária (já que estão contidos em peptídeos diferentes após a proteólise), os resíduos reativos se encontram próximos nas estruturas superiores devido ao enovelamento característico da proteína alvo. No caso de complexos proteicos, as espécies intermoleculares podem ainda fornecer informações sobre a interface de contato entre as proteínas, caso cada extremidade do ALC tenha reagido com um dos monômeros.



**Figura 9.** Representação dos tipos de espécies formadas em experimentos de ligação cruzada. Em A) uma representação dessas espécies após a proteólise enzimática da proteína (ou complexo) alvo. Em B) uma representação dessas espécies ligadas na superfície da proteína antes da proteólise enzimática.

### 1.5. Desafios e avanços na área de ligação cruzada

As amostras de proteínas alvo submetidas a experimentos de ligação cruzada são complexas, e os peptídeos modificados pelos ALCs se encontram em

quantidades subestequiométricas em relação aos peptídeos comuns. Dessa forma, os dados obtidos nesses experimentos devem ser adquiridos em equipamentos com alta velocidade de aquisição e, obviamente, alta resolução/exatidão, para que se tenha maior confiabilidade na identificação das espécies modificadas. Essas necessidades instrumentais acabaram limitando o amplo uso da técnica no início dos anos 2000, algo que tem sido resolvido com a disponibilidade comercial de plataformas Q-TOF que adquirem até 20 espectros de íons produtos por segundo<sup>45</sup> e plataformas como Q-Orbitrap com até 20 espectros por segundo com alta resolução<sup>46</sup>.

Outro grande desafio enfrentado pela técnica de XL-MS foi o desenvolvimento de softwares para a identificação das espécies modificadas pelos ALCs. Inicialmente, as buscas eram feitas de forma manual, levando à identificação de um número bastante reduzido de espécies de ligação cruzada em meio à enorme quantidade de espectros obtidos nesses experimentos. No final da década de 2000, algumas versões adaptadas de softwares de proteômica convencional começaram a surgir, embora ainda muito primitivos devido à falta de uma interface amigável e intuitiva e baixa sensibilidade. Esses softwares apenas forneciam uma tabela de possíveis candidatos, e a validação dessas espécies deveria ser feita manualmente analisando-se os espectros. Em 2014, juntamente com o grupo do Dr. Paulo Costa Carvalho (FIOCRUZ-PR), nosso grupo de pesquisa desenvolveu um software direcionado para a identificação das espécies de ligação cruzada, com interface intuitiva e visualização gráfica dos espectros de massa, chamado SIM-XL (*Spectrum Identification Machine for Cross-Linked peptides*)<sup>47</sup>, o que acelerou e simplificou enormemente a validação dos dados experimentais e obtenção das restrições de distância por XL-MS. Juntamente ao SIM-XL, alguns outros softwares com o mesmo propósito foram lançados pela comunidade, com maior destaque ao p-Link<sup>48</sup>, MS-Studio<sup>49</sup> e StavroX<sup>50</sup>, que recebem atualizações com frequência, indicando forte desenvolvimento nessa área com o objetivo de ampliar cada vez mais a utilização da metodologia de XL-MS.

Além dos campos de instrumentação e software, há paralelamente (e ainda mais ativamente) o constante desenvolvimento de novos ALCs, com diferentes especificidades e características que tornam os experimentos de XL-MS cada vez mais versáteis. Uma metodologia bastante explorada é o uso de reagentes isotopicamente marcados, na qual uma mistura 1:1 dos ALCs leves e pesados é utilizada no experimento de ligação cruzada. As amostras são posteriormente

analisadas e deve-se esperar que as espécies de ligação cruzada formadas nos experimentos apareçam como sinais duplicados nos espectros de MS, facilitando sua identificação<sup>51,52</sup>. A desvantagem dessa metodologia é que, além de tornar a mistura mais complexa, a intensidade do sinal de cada espécie cai pela metade, já que agora metade delas possui massa diferente.

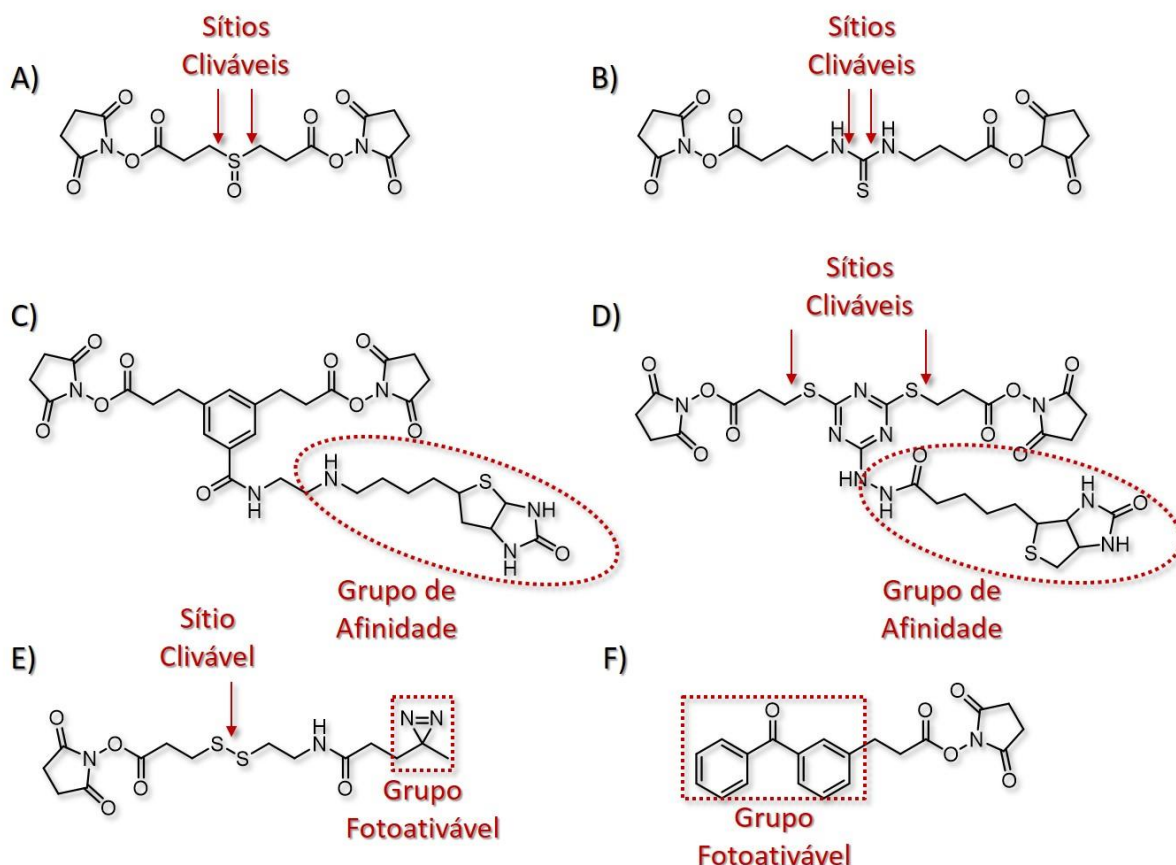
Outra metodologia bastante explorada é a utilização de ALCs bifuncionais com sítios de clivagem na cadeia espaçadora (Figura 10A e Figura 10B)<sup>52,53</sup>. No uso desses ALCs, ocorre a fragmentação preferencial desses sítios de clivagem nos experimentos de MS<sup>2</sup>, formando fragmentos que atuam como modificações pós-traducionais de massa conhecida nos peptídeos que reagiram, e isso facilita sua identificação, principalmente pelos softwares de proteômica convencional que estavam disponíveis na época do desenvolvimento dessa metodologia. Nessa abordagem, a fragmentação da cadeia peptídica deve ocorrer em experimentos de MS<sup>3</sup>, e isso consiste em sua grande desvantagem: além da necessidade de instrumentos capazes de realizar experimentos de MS<sup>3</sup>, os espectros obtidos nesses experimentos são ruidosos e geralmente de baixa qualidade devido à menor intensidade dos precursores.

Uma estratégia diferente, desenvolvida para tentar contornar o fato de que a concentração das espécies de ligação cruzada é muito menor em relação a dos peptídeos não modificados, é a utilização de ALCs que possuem um grupo de afinidade<sup>51,54</sup>, geralmente biotina, (Figura 10C e Figura 10D), de modo a ser possível o enriquecimento dessas espécies após a digestão enzimática. A desvantagem desse método está relacionada à labilidade das extremidades do ALC (que são ésteres derivados de N-hidroxysuccinimida, NHS). Na presença de um grupo com caráter nucleofílico na molécula do ALC, ocorre o ataque desse grupo na extremidade, impossibilitando a ligação das duas extremidades em resíduos diferentes da proteína. Além disso, no caso de biotina, a fragmentação das espécies enriquecidas é normalmente mais difícil de ser interpretada, prejudicando a identificação das mesmas.

Um fato interessante é que todos os ALCs mencionados até o momento são homobifuncionais e reativos frente a resíduos de lisina. Em uma tentativa de aumentar o número de resíduos que podem ser reativos nos experimentos de ligação cruzada, foram desenvolvidos reagentes heterobifuncionais fotoativáveis<sup>55,56</sup> (Figura 10E e Figura 10F), nos quais uma extremidade é reativa frente a lisinas e, uma vez



que o ALC está ancorado na superfície da proteína por essa extremidade, é incidida radiação de comprimento de onda específico para ativar a outra extremidade. Esse grupo é reativo de maneira inespecífica, de modo que pode reagir com qualquer outro resíduo que esteja dentro do alcance da cadeia espaçadora. Embora esses ALCs de fato aumentem o número de espécies formadas nos experimentos de ligação cruzada, adiciona-se uma dificuldade ainda maior na identificação das espécies que foram formadas, já que o número de combinações possíveis se torna muito grande.



**Figura 10.** Estrutura química de alguns ALCs. A e B) ALCs com sítios cliváveis; C e D) ALCs com grupos de afinidade; e E e F) ALCs heterobifuncionais fotoativáveis.

As diferentes estratégias mencionadas trazem consigo certas vantagens e desvantagens, e mesmo com seu desenvolvimento algumas limitações seguem impedindo o uso mais abrangente da técnica de ligação cruzada associada à espectrometria de massas para resolução estrutural de proteínas e complexos; sendo as mais importantes: 1) a obtenção de um pequeno número de restrições de distância entre os resíduos, que fornecem informações limitadas para a resolução estrutural de proteínas através da modelagem computacional; 2) o baixo número de resíduos reativos frente aos ALCs utilizados como padrão na área; 3) a estequiometria



desfavorável das espécies de ligação cruzada formadas, que estão em quantidade muito inferior aos peptídeos convencionais da proteína-alvo, dificultando sua detecção e identificação; e 4) a identificação de falsos positivos pelos softwares de análise dos dados, que fornecem informações falsas a respeito da distância entre os resíduos, prejudicando enormemente a geração de modelos confiáveis e condizentes com a estrutura nativa da proteína.<sup>57</sup>

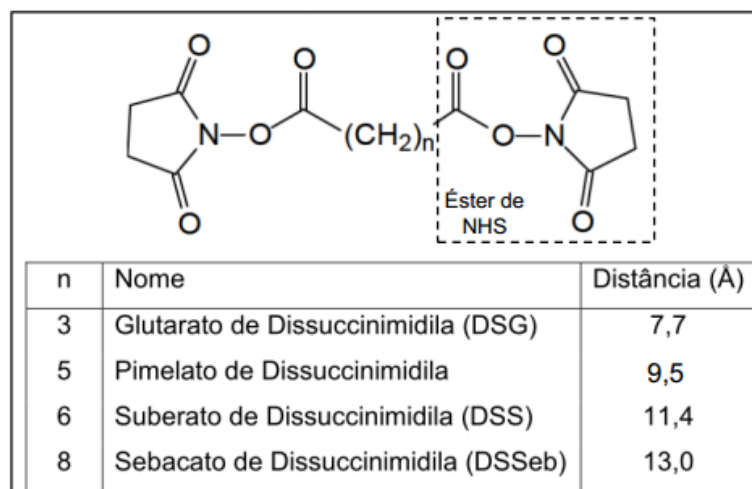
Como mencionado, apesar do intenso e contínuo desenvolvimento no campo de novos ALCs, a maior parte das soluções propostas trazem consigo algumas desvantagens, além de, normalmente, serem moléculas de difícil obtenção. Por essas razões, os ALCs mais amplamente utilizados, considerados padrões em experimentos de XL-MS, são os derivados de ésteres de N-hidroxisuccinimida (NHS), que serão discutidos com maior detalhe a seguir.

#### **1.6. Uso de derivados de ésteres de NHS como ALCs e suas limitações**

Dentre as características necessárias e desejadas em um ALC para experimentos de XL-MS, está a capacidade de reagir com proteínas ou complexos proteicos em condições próximas à nativa, *i.e.* condições brandas como pH próximo de 7 e solução 100% aquosa. A temperatura da reação também deve ser a mais baixa possível e a reação deve ocorrer de forma rápida (dentro de minutos, preferencialmente), para que as informações de distância obtidas sejam amostradas da proteína em sua forma nativa e a mesma não sofra degradação no meio reacional devido a longos períodos do experimento e/ou temperaturas elevadas. Obedecendo-se a todas essas condições, espera-se ainda que a reação tenha bom rendimento, ou seja, que o máximo de restrições possíveis sejam obtidas em um único experimento.

Essa série de requisitos, como seria de se esperar, restringe enormemente o número de possíveis candidatos viáveis a serem usados como ALCs. Dentre os reagentes comerciais disponíveis para este fim, os mais utilizados são os derivados de ésteres de NHS (Figura 11), com maior destaque ao suberato de dissuccinimidila (DSS), considerado o ALC padrão para experimentos de XL-MS. Esses reagentes são reativos preferencialmente frente a grupos nucleofílicos como aminas primárias (resíduos de lisina e N-terminal da proteína), podendo reagir também com grupos álcool e tiol (resíduos de serina, treonina e cisteína), em menor escala<sup>29</sup>. Seu uso como ALC é preferido devido a sua fácil obtenção, bastando reagir um haleto de acila

com o NHS, e aos rendimentos razoáveis das reações nas condições brandas citadas anteriormente, exceto a temperatura como será discutido posteriormente.



**Figura 11.** Série de agentes de ligação cruzada homólogos derivados de ésteres de NHS com diferentes tamanhos de cadeia alifática espaçadora. O ALC mais utilizado, DSS, tem  $n=6$  e uma cadeia espaçadora que mede 11,4 Å.

Apesar de utilizado como padrão na área de ligação cruzada, há uma série de limitações enfrentadas na realização de experimentos de XL-MS utilizando o DSS como ALC. Uma dessas desvantagens diz respeito a sua baixa solubilidade em meio aquoso. Como alternativa, o DSS é primeiramente solubilizado em dimetilformamida (DMF) anidra, e dessa solução é aliquoteada a quantidade de ALC necessária para a reação (em geral, 100 vezes de excesso molar em relação à proteína). O problema existente neste método é que o volume de DMF não deve ultrapassar 10% do volume total da reação (ficando idealmente abaixo de 5%), pois uma mudança expressiva de polaridade no meio poderia causar mudanças conformacionais na estrutura da proteína e levar à obtenção de dados de distância entre resíduos que não possuem correlação com sua estrutura nativa em solução. Há disponível comercialmente um derivado sulfonado do DSS (chamado de BS3), no qual um dos hidrogênios do ciclo do NHS é substituído por  $-SO_3$ , sendo mais solúvel em meio aquoso. No entanto, o valor desse reagente chega a ser entre 20 e 50 vezes maior que o DSS.

Uma outra limitação desse ALC está relacionada à presença dos grupos NHS nas extremidades da molécula. Por serem muito lábeis, esses grupos sofrem hidrólise facilmente em solução, de modo que é favorecida a formação de espécies do tipo *dead-end*. Essas espécies, no caso dos experimentos com DSS, não são informativas a respeito da estrutura e acessibilidade dos resíduos reativos, pois não

há como afirmar as espécies do tipo *dead-end* são formadas pela ausência de resíduos reativos dentro do alcance da cadeia espaçadora ou simplesmente pela hidrólise de uma das extremidades do ALC antes que ela pudesse reagir com outro resíduo. Além disso, a labilidade dos grupos NHS impede que outros grupos funcionais que pudessem ser úteis nas análises por MS (como grupos de afinidade ou etiquetas de defeito de massa, por exemplo) possam estar presentes na molécula do ALC.

Por fim, as reações de ligação cruzada utilizando o DSS como ALC são muito sensíveis à temperatura, de modo que o rendimento das mesmas cai drasticamente em temperaturas menores. Essa limitação é particularmente importante pois, além da óbvia implicação de que seria obtido um conjunto menor de restrições de distância entre resíduos, há também o fato de que um grande número de proteínas pode sofrer desenovelamento ou agregação em temperaturas mais altas, e por isso não são passíveis de serem estudadas por XL-MS utilizando-se o DSS. Outro ponto importante a ser salientado é que quanto mais baixa a temperatura da reação menor a energia interna das moléculas de proteína, ou seja, menos estados conformacionais são populados, de modo que um maior número de moléculas deve se encontrar no estado de menor energia. Isso faz com que os dados de restrição de distância entre resíduos ligados pelo ALC sejam mais representativos, guiando o processo de modelagem computacional para a geração de melhores modelos, mais convergentes com a estrutura nativa da proteína.

Diante das limitações apresentadas pelo uso de DSS (e derivados de ésteres de NHS em geral), neste trabalho foi desenvolvida uma metodologia alternativa utilizando o ácido adípico (ácido hexanodioico) como ALC. Esse reagente é reativo também frente aos grupos nucleofílicos (lisinas, serinas, N-terminal, etc.), mas a metodologia desenvolvida ainda permitiu que resíduos de ácido aspártico e ácido glutâmico da proteína formem espécies de ligação cruzada com os esses grupos nucleofílicos, multiplexando as reações de ligação cruzada em um mesmo experimento e se apresentando como uma alternativa vantajosa frente ao uso dos ésteres de NHS.

## **2. OBJETIVOS**

### **2.1. Objetivo Geral**

Desenvolver uma nova metodologia de ligação cruzada associada à espectrometria de massas envolvendo a utilização de diácidos carboxílicos como agente de ligação cruzada frente a resíduos nucleofílicos presentes em proteínas.

### **2.2. Objetivos Específicos**

- Avaliar a capacidade de EDC/HOBt na ativação de ácidos dicarboxílicos;
- Avaliar a reatividade de ácidos dicarboxílicos ativados com EDC/HOBt frente a resíduos nucleofílicos
- Otimizar a reação de ligação cruzada e avaliar as variáveis reacionais;
- Comparar a eficiência dessa nova reação de ligação cruzada com reagentes comerciais baseados em ésteres de NHS.

### 3. PROCEDIMENTO EXPERIMENTAL

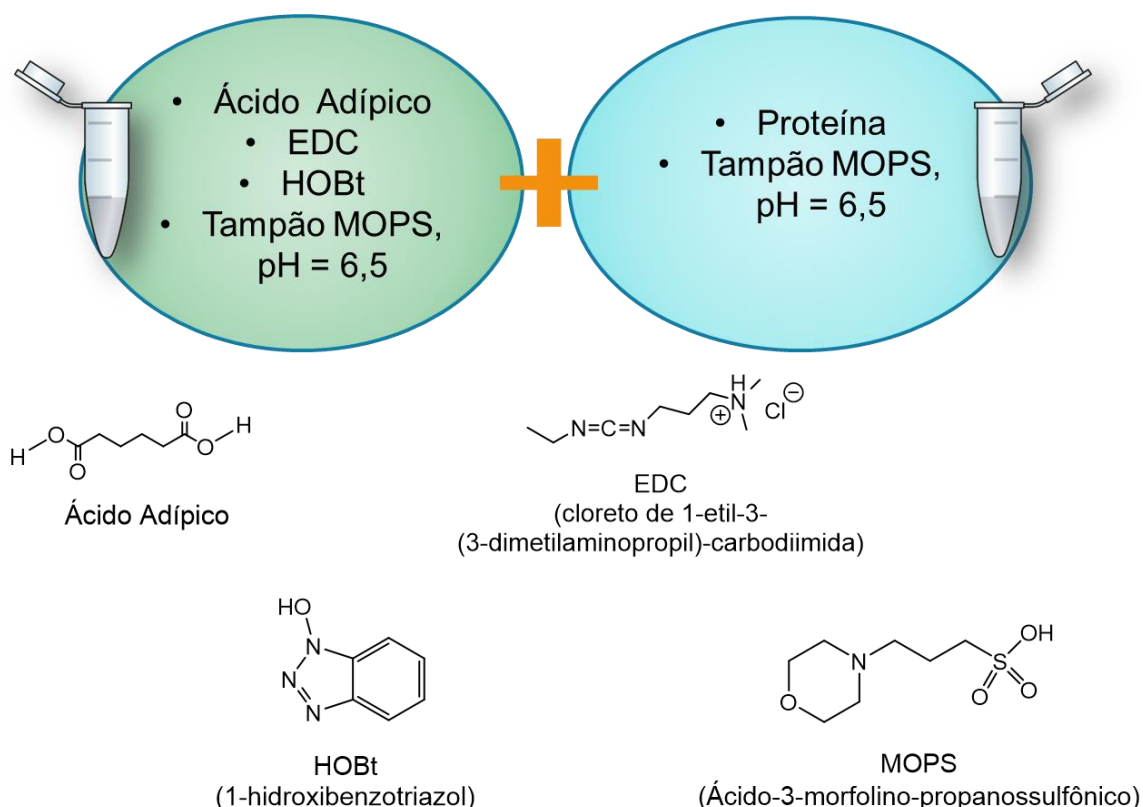
Para o desenvolvimento metodológico do projeto, foi utilizada a mioglobina (de coração de cavalo, *Sigma Aldrich*) como proteína para os testes, sendo ela largamente utilizada pela comunidade científica para esses fins, visto que é facilmente obtida, relativamente barata, possui vasto acervo de informações disponível, além de estrutura terciária resolvida (entrada no PDB: 1WLA).

Para avaliar a robustez da metodologia proposta, as reações foram realizadas na proteína citocromo c (de coração de cavalo, *Sigma Aldrich*), também muito empregada como padrão no desenvolvimento de métodos, com estrutura também já resolvida (entrada no PDB: 1HRC), utilizando-se as condições otimizadas obtidas através dos testes da mioglobina. As amostras de citocromo c não passaram por proteólise enzimática e análise por LC-MS/MS

#### 3.1. Ensaios de ligação cruzada utilizando o ácido adípico como ALC

As reações utilizando o ácido adípico foram sempre divididas em duas etapas, uma de ativação dos grupos ácidos do ALC, e outra de reação entre o ALC e a proteína. Em um frasco, a mioglobina foi dissolvida em tampão MOPS (ácido 3-(n-morfolino)propanossulfônico) 300 mM, em pH = 6,5 (ajustado com solução 1 M de hidróxido de sódio). Em outro frasco, ácido adípico, cloreto de 1-etil-3-(3-dimetilaminopropil)-carbodiimida (EDC) e o aditivo 1-hidroxibenzotriazol (HOBt), foram solubilizados no mesmo tampão, em mesmo pH. Este pH foi utilizado pois, além de ser suficientemente próximo de neutro para que não ocorram mudanças significativas na estrutura da proteína, este é o valor ótimo para a ativação dos grupos ácidos com EDC e HOBt e posterior ataque dos resíduos nucleofílicos, conforme estabelecido em recente projeto de doutorado desenvolvido no nosso grupo de pesquisa. Decorrido certo tempo, a solução contendo o ALC (agora ativado) foi vertida na solução contendo a proteína para que o ocorresse a reação.

Foram testadas diferentes concentrações dos reagentes, bem como a ocorrência de reação na ausência de um ou mais deles. Foram também testadas diferentes temperaturas (5, 15 e 25 °C) de reação, além de diferentes períodos de ativação (5, 10, 20, 30 e 60 minutos) dos grupos ácidos em solução contendo apenas ácido adípico, EDC e HOBt no tampão (somente sem a proteína). A Figura 12 ilustra o procedimento e as moléculas utilizadas nesta etapa.



**Figura 12.** Representação esquemática da realização do experimento de ligação cruzada utilizando o ácido adípico como ALC. Na parte de baixo da figura, estão representadas as estruturas das moléculas utilizadas nessas reações.

### 3.2. Ensaios de ligação cruzada utilizando o DSS como ALC

Para efeito de comparação com a metodologia utilizada como padrão em experimentos de ligação cruzada, foram realizadas reações tanto na mioglobina quanto no citocromo c empregando como ALC o DSS (suberato de dissuccinimidila, *Sigma Aldrich*). Nessas reações, a proteína foi dissolvida em tampão MOPS, pH = 6,5, e o DSS foi solubilizado em DMF, formando uma solução de concentração final 27 mM. O volume aliquotado dessa solução do ALC foi tal que o excesso molar de DSS correspondesse a 100 vezes em relação a proteína, e o volume final da reação foi ajustado anteriormente de modo a garantir que com a adição da solução de DSS, o volume de DMF não ultrapassasse 5% do volume total. As reações foram realizadas a 5, 15 e 27 °C, sendo essa última a temperatura ótima da reação com esse ALC.

### 3.3. Análise de proteína intacta modificada pelo ALC por MS

Após o término das reações de ligação cruzada, as amostras foram dessalinizadas através de cartuchos de extração em fase sólida (SPE) Oasis HLB (*Waters Co.*), que retém as macromoléculas mas não os outros reagentes orgânicos. Os produtos de reação presos no cartucho foram lavados com 1 mL de água por 10 vezes e com 1 mL de uma mistura água:acetonitrila 80:20 (v/v). A proteína modificada com o ALC foi eluída do cartucho com 600  $\mu$ L de uma mistura água:acetonitrila 50:50 (v/v) contendo 0,1% em volume de ácido fórmico, para ionização da proteína ao ser injetada no espectrômetro de massas. Uma alíquota de 100  $\mu$ L foi retirada para injeção no espectrômetro de massas, enquanto que os outros 500  $\mu$ L foram submetidos à evaporação da acetonitrila e a solução resultante separada para posterior proteólise enzimática.

As amostras de proteína intacta foram analisadas por infusão direta com fluxo de 100  $\mu$ L/min em um espectrômetro de massas Waters Synapt G1 HDMS, com ionização por *electrospray* (ESI). Foram injetados 100  $\mu$ L de cada fração (aproximadamente 5 nmol de proteína por injeção) na fonte de ESI do espectrômetro a um fluxo de 50  $\mu$ L min<sup>-1</sup>. Os espectros foram adquiridos no modo V, na faixa de *m/z* 800 a 3000. A energia de colisão aplicada foi definida em 3 eV, apenas para garantir que aglomerados proteicos fossem desfeitos.

### 3.4. Proteólise enzimática

As amostras de mioglobina intacta foram primeiramente submetidas a uma etapa de desnaturação. Para isso, foi adicionada uma solução de guanidina hidrocloreada a proteína de modo que a concentração final de guanidina no frasco fosse 6M. Essa solução foi aquecida a 65 °C por 30 minutos. O excesso de guanidina foi removido submetendo a solução à troca de tampão utilizando um filtro de centrífuga Centricon de 5 kDa (*Amicon*), e a proteína foi ressuspensa em tampão bicarbonato de sódio 100 mM, pH=8. As proteínas foram submetidas à dupla digestão enzimática com tripsina (*sequencing grade modified porcine pancreas, Promega*) na proporção de 1:30 (m/m, tripsina/proteína), a 37 °C, por 32h (uma adição de tripsina a cada 16h). Após a digestão, as soluções dos peptídeos foram dessalinizadas utilizando-se cartuchos de extração em fase sólida (SPE) Oasis HLB (*Waters Co.*) e efetuando-se dez lavagens com 1 mL de água cada, e os peptídeos foram eluídos com 500  $\mu$ L de uma mistura água:acetonitrila 70:30 (v/v).

### 3.5. Análise dos peptídeos modificados pelo ALC por MS

As análises dos peptídeos obtidos após a proteólise enzimática da mioglobina foram realizadas por UPLC-MS e UPLC-MS/MS, utilizando-se utilizando um cromatógrafo líquido de nanofluxo Easy-nLC 1200 (*ThermoFisher Scientific*), acoplado a um espectrômetro de massas Orbitrap Fusion Lumos Tribrid (*ThermoFisher Scientific*), localizado no Instituto de Química da Universidade de São Paulo (IQ/USP, São Paulo).

Foram injetados aproximadamente 10 picomols de cada uma das amostras, e as corridas cromatográficas foram de 30 minutos. As rampas de gradientes foram de 5 a 30% (v/v) acetonitrila em água + 0,1% de ácido fórmico entre 0 a 20 minutos, 30 a 90% (v/v) acetonitrila em água + 0,1% de ácido fórmico entre 20 e 25 minutos, e manteve-se 90% (v/v) acetonitrila em água + 0,1% de ácido fórmico até o fim dos 30 minutos. Os dados de espectrometria de massas foram adquiridos de modo contínua durante toda a cromatografia. Cada espectro de MS foi adquirido no Orbitrap, com resolução de 60000 ( $m/z$  400) e na faixa de  $m/z$  entre 375 e 2000. O tempo máximo de acúmulo de íons foi de 50 ms, e valor alvo de  $4 \times 10^5$  íons. Os fragmentos foram gerados em ciclos de 3 segundos, tomando os íons precursores mais intensos nesse intervalo. Os espectros de MS<sup>2</sup> foram adquiridos de duas maneiras. Íons precursores de carga +2 e +3 foram fragmentados e analisados no ion trap linear, com tempo de injeção máximo de 70 ms e valor alvo de  $1 \times 10^4$  íons. Os precursores de cargas mais altas foram fragmentados por CID e analisados no Orbitrap, com resolução 30000, tempo máximo de injeção de 300 ms e valor alvo de  $1 \times 10^5$  íons.

### 3.6. Tratamento dos dados e identificação das espécies de ligação cruzada

Os espectros de massas das proteínas intactas obtidos no espectrômetro de massas Waters Synapt G1 HDMS foram deconvoluídos utilizando-se o software MaxEnt 1, módulo embutido na suíte MassLynx (*Waters Co.*), obtendo-se os sinais monocarregados das espécies (M+H)<sup>+</sup>. Os parâmetros de faixa de  $m/z$  foram fixados entre 16500 e 20500 para a mioglobina e entre 12000 e 16000 para o citocromo c; a resolução foi fixada em 1 Da/canal e a largura a meia altura dos sinais foi analisada caso a caso.

Já os dados dos peptídeos da mioglobina advindos da digestão, obtidos no espectrômetro de massas Orbitrap Fusion Lumos Tribrid, foram submetidos a duas etapas de análise. Primeiramente, foi analisada a cobertura de sequência da proteína



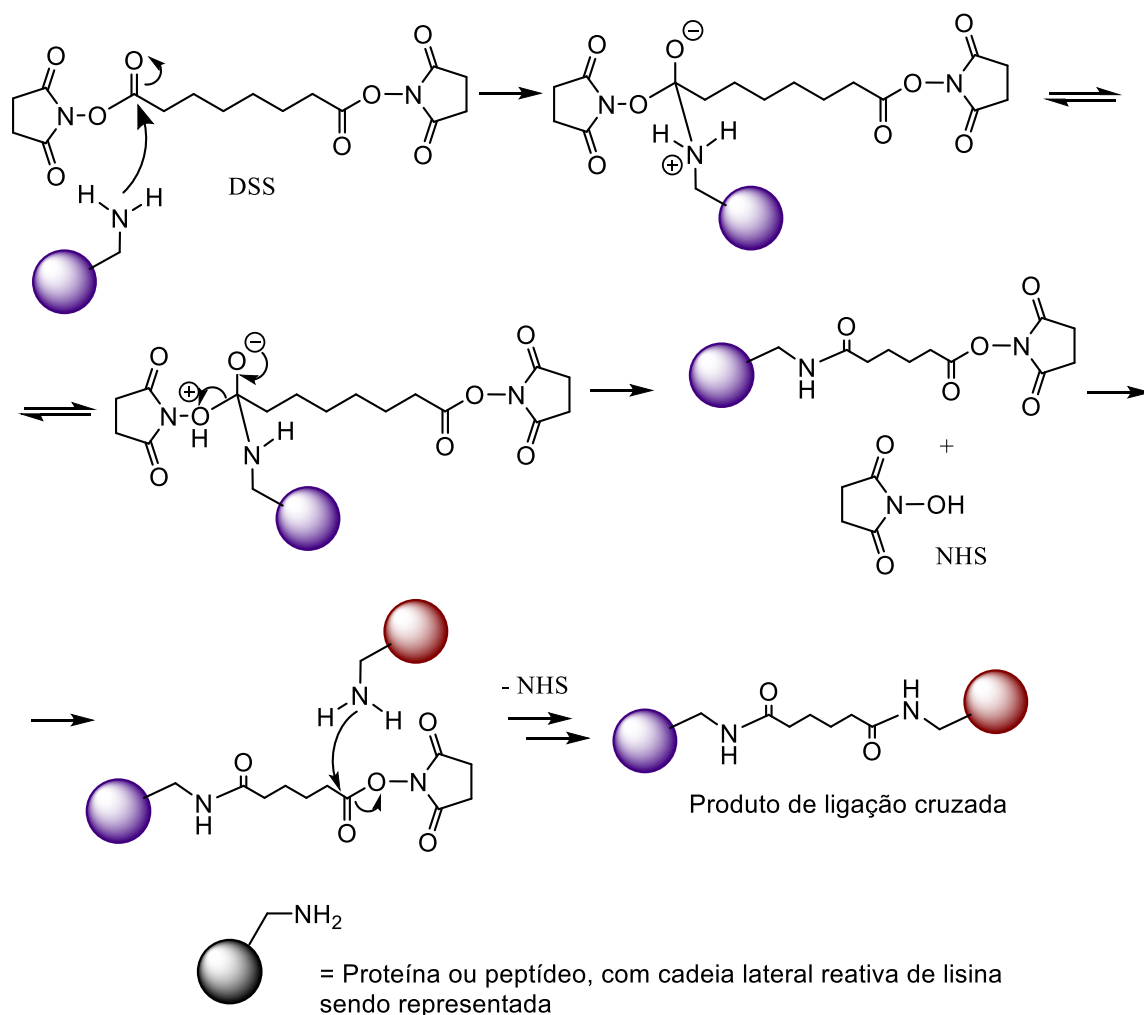
através do software Mascot (*Matrix Science*), utilizando como parâmetros de busca o banco de dados SwissProt, 15 ppm de erro no modo MS e 0,8 Da no modo MS/MS, tripsina como enzima da proteólise e três sítios de clivagem não digeridos. Foram adicionadas como modificações variáveis a oxidação de metionina e as adições de DSS (156,08 Da) e ácido adípico (128,05 Da) para identificação das espécies do tipo *dead-end*. Os dados brutos das corridas foram convertidos para o formato MGF, utilizado pelo Mascot, através da ferramenta de conversão de arquivos do SIM-XL.

A segunda análise realizada foi a identificação das espécies de ligação cruzada, e para isso foi utilizado o software SIM-XL. Os parâmetros de busca utilizados foram 20 ppm de erro para as massas tanto do precursor quanto do fragmento, e o *cross-linker* (ALC) foi escolhido de acordo com o experimento, sendo que as massas das modificações por ácido adípico (128,05 Da para *dead-end* e 110,04 Da para ligação cruzada) foram adicionadas manualmente. Foram permitidos até três sítios de clivagem não digeridos por tripsina e a oxidação de metionina foi adicionada como modificação variável. Foram validadas e aceitas espécies de ligação cruzada que apresentassem *score* interpeptídico acima de 3,5 e intrapeptídico acima de 2,5, além de cada uma das espécies terem *spectral count* de no mínimo 2.

## 4. RESULTADOS E DISCUSSÃO

### 4.1. Uso do DSS como ALC reativo frente a resíduos nucleofílicos e possíveis alternativas

Ésteres de NHS reagem principalmente com aminas e álcoois primários (presentes em resíduos de lisina e serina) formando amidas e ésteres, respectivamente, por um mecanismo simples de adição-eliminação nucleofílica, característico de reações de derivados de ácidos carboxílicos<sup>58</sup> (Figura 13). O nucleófilo, representado na figura por uma lisina, ataca o carbono da carbonila do DSS, levando à formação de um intermediário tetraédrico (com o carbono em hibridização  $sp^3$ ). No equilíbrio de transferência de prótons o oxigênio ligado ao nitrogênio é protonado e todo o grupo é eliminado como NHS ao se reestabelecer a hibridização  $sp^2$  ao carbono da carbonila. Este grupo é eliminado preferencialmente devido à formação de uma amida (mais estável), e o NHS ser um bom grupo de saída. Caso exista na proteína outro resíduo nucleofílico espacialmente próximo, ou seja, dentro do alcance da cadeia espaçadora, a outra extremidade do DSS sofre a mesma reação e é formado um produto de ligação cruzada. Caso não exista um resíduo espacialmente próximo será formada uma espécie do tipo *dead-end* como resultado da hidrólise da outra extremidade do ALC.



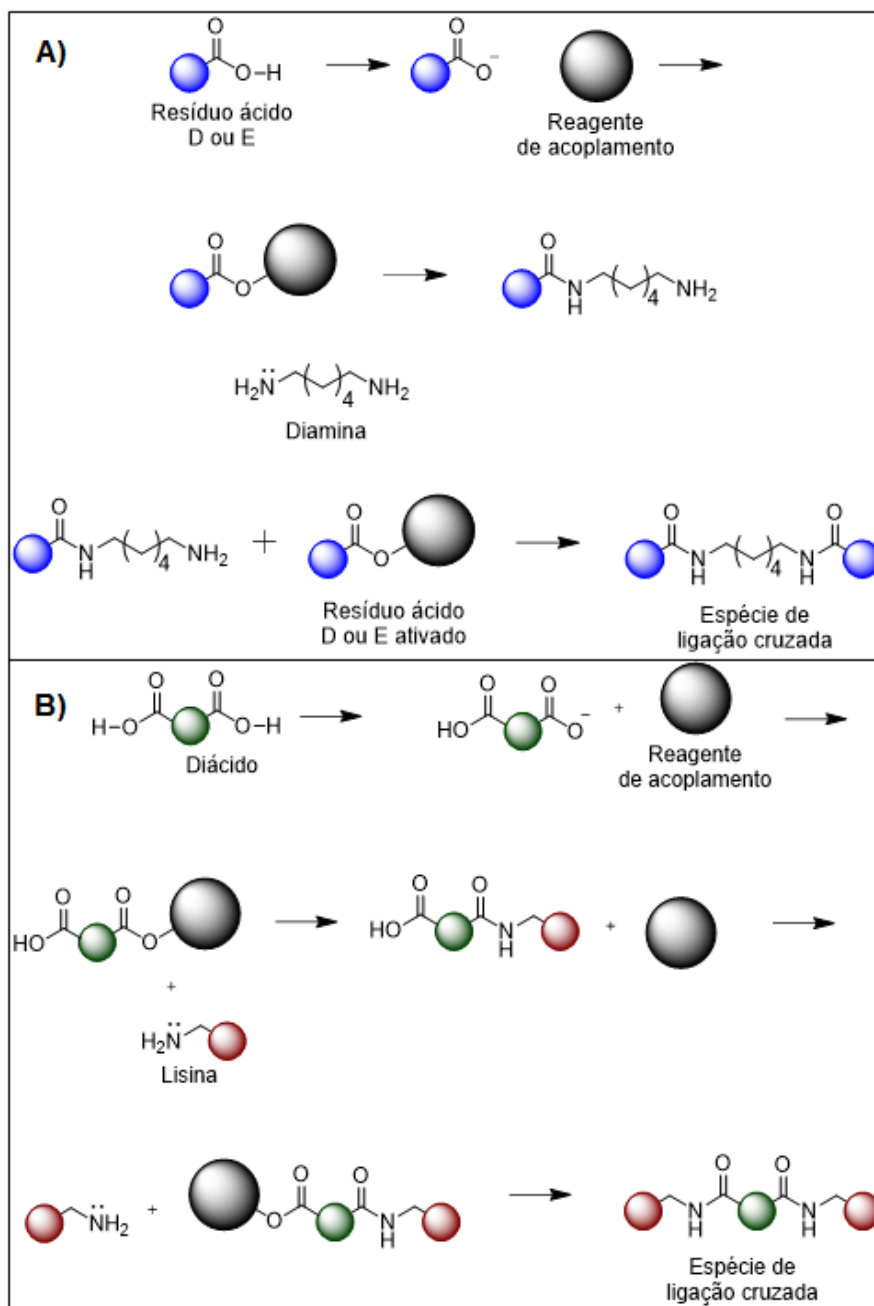
**Figura 13.** Representação do mecanismo de formação de uma espécie de ligação cruzada utilizando o DSS como ALC sendo atacado por um grupo nucleofílico de um resíduo de lisina.

Como as reações dos resíduos nucleofílicos com o DSS seguem um mecanismo semelhante aos de adição-eliminação nucleofílica de derivados de ácidos carboxílicos, uma alternativa natural e até intuitiva para substituir esse ALC seria a utilização dos próprios derivados de ácidos como ALCs. De acordo com a ordem de reatividade desses compostos<sup>58</sup>, haletos de acila seriam os mais reativos e possíveis candidatos a ALCs, mas sua reatividade é tão elevada que, no meio aquoso em que os experimentos de ligação cruzada são efetuados, a taxa de hidrólise e formação do respectivo ácido carboxílico seria muito alta, além da possível alteração do pH devido à eliminação de ácido clorídrico no meio. O mesmo raciocínio pode ser aplicado para os anidridos de ácidos, logo abaixo dos haletos na escala de reatividade. Continuando nessa escala estão os ácidos carboxílicos, que não possuem a desvantagem da hidrólise pois são regenerados quando atacados pela água, mas sua reatividade frente aos resíduos nucleofílicos das proteínas é muito menor, a ponto de

praticamente não reagirem nas condições brandas dos experimentos de ligação cruzada. No entanto, há diversos trabalhos reportados na literatura nos quais se utilizam reagentes de acoplamento, que ativam os grupos ácidos e assim aumentam sua reatividade<sup>59,60</sup>. Dessa forma, a classe de ácidos carboxílicos, mais precisamente diácidos (para que ocorra a ligação cruzada entre dois resíduos), se tornam candidatos interessantes a possíveis ALCs.

#### **4.2. Dupla ativação de grupos ácidos e uso do ácido adípico como ALC**

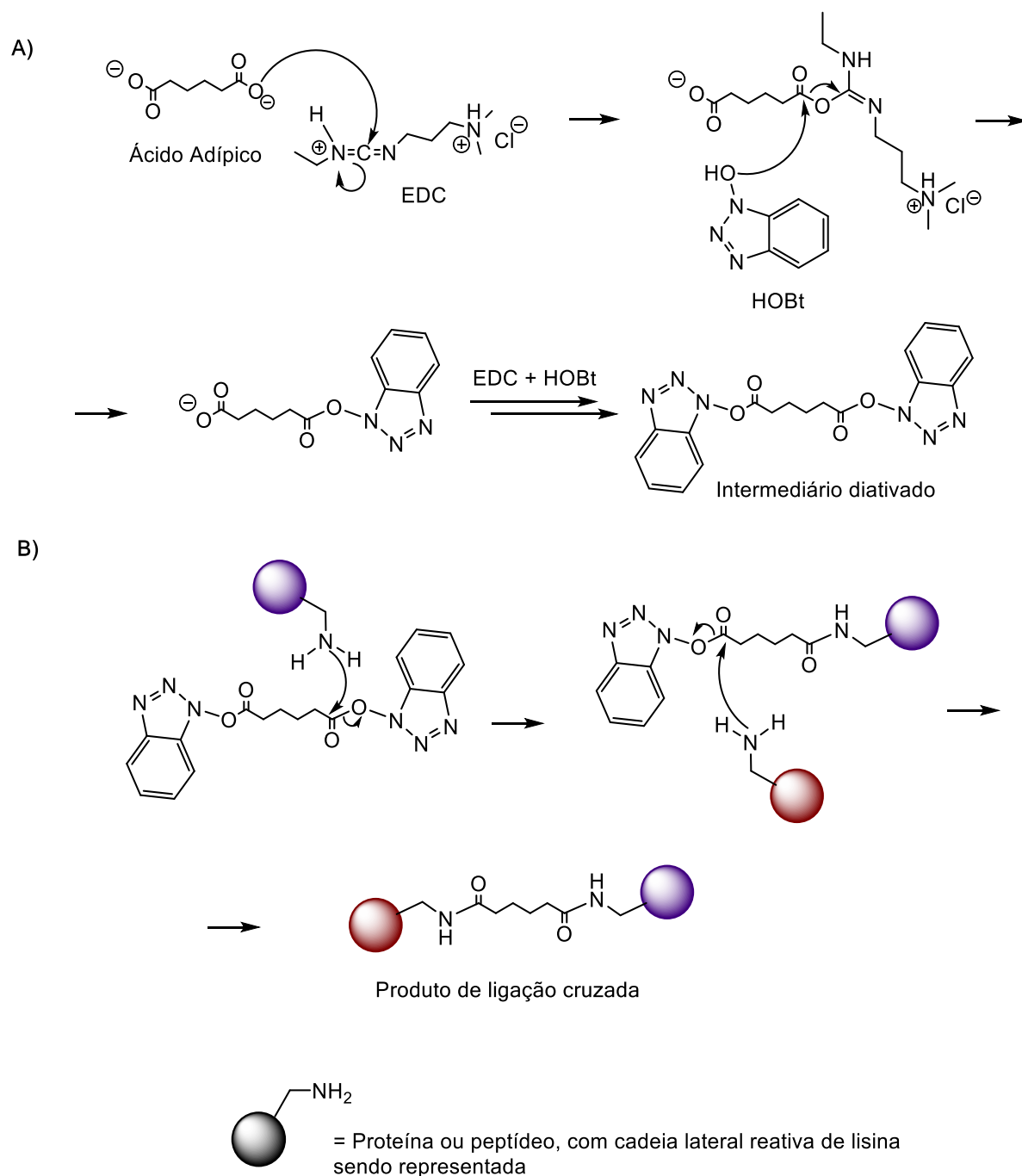
A ideia por trás de ativar um grupo ácido é aumentar sua reatividade, adicionando à molécula um bom grupo de saída. Esse grupo de saída é facilmente eliminado quando um nucleófilo ataca a carbonila da molécula ativada. Em trabalho recente de nosso grupo de pesquisa foi desenvolvida uma nova química de ligação cruzada, na qual resíduos eletrofílicos de ácido aspártico e glutâmico são ativados por reagentes de acoplamento, que permitem o uso de diaminas como ALCs atacando as espécies ativadas e formando produtos de ligação cruzada (Figura 14A). A partir desses conceitos, foi possível adaptar essa abordagem de ligação cruzada de modo a se utilizar um diácido como ALC. Os grupos ácidos do ALC puderam então ser ativados pelos reagentes de acoplamento e se ficaram passíveis de serem atacados pelas cadeias laterais nucleofílicas de lisinas e serinas (Figura 14B).



**Figura 14.** Representação da etapa de ativação dos grupos ácidos nas duas metodologias desenvolvidas no grupo de pesquisa. A) ativação dos grupos ácidos dos resíduos de ácido aspártico de glutâmico e uso da 1,6-hexanodiamina como ALC, formando espécies de ligação cruzada entre esses resíduos eletrofílicos; B) ativação dos grupos ácidos de um diácido usado como ALC e a reação da espécie ativada com resíduos de lisina, formando espécies de ligação cruzada entre esses resíduos nucleofílicos.

Assim como na metodologia da 1,6-hexanodiamina, foi empregada uma dupla ativação com os reagentes de acoplamento EDC e HOBt, de uso bastante difundido em experimentos envolvendo proteínas e síntese de aminoácidos<sup>59,61</sup>. O

mecanismo da dupla ativação e da reação das espécies ativadas com resíduos nucleofílicos são mostrados na Figura 15.

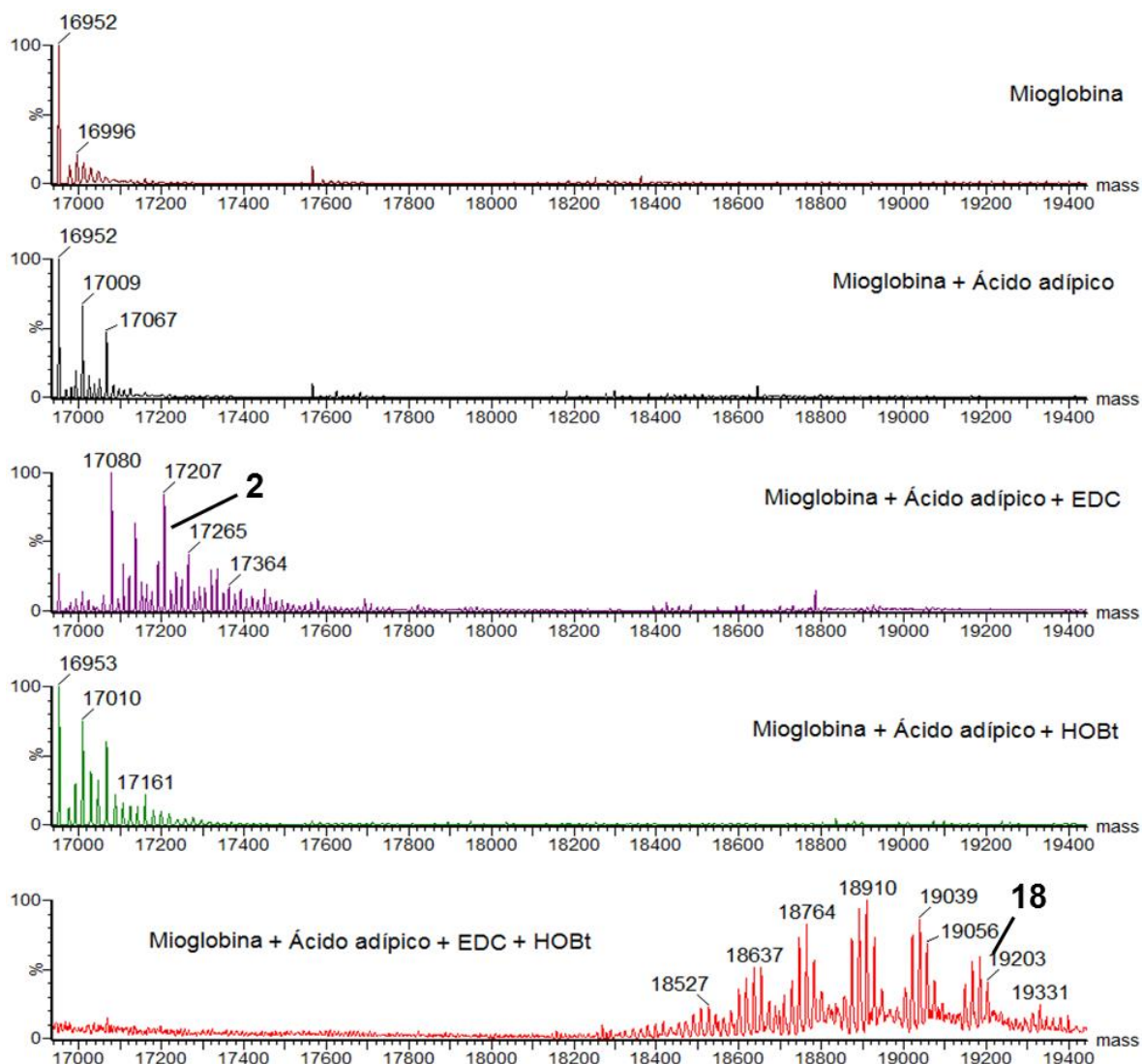


**Figura 15.** Representação do mecanismo de formação de produtos de ligação cruzada utilizando ácido adípico como ALC. A) mecanismo da dupla ativação do ácido adípico com EDC e HOBT; B) mecanismo da reação da espécie ativada com resíduos de lisina para formação das espécies de ligação cruzada.

Em pH fisiológico, os dois grupos ácidos do ácido adípico se encontram majoritariamente desprotonados ( $\text{pK}_a$  4,4 e 5,4), enquanto que o EDC se encontra em

sua forma protonada. O ânion carboxilato ataca o carbono mais deficiente em elétrons do EDC, formando a primeira espécie ativada. Essa espécie é então atacada pelo HOBt, eliminando o EDC como um derivado de ureia (excelente grupo de saída). A espécie duplamente ativada agora pode ser atacada mais eficientemente pelos resíduos nucleofílicos da proteína, formando os produtos de ligação cruzada. A eliminação do HOBt se dá pelo mesmo mecanismo, e é favorecida pois, devido ao caráter altamente ácido ( $pK_a = 4,6$ ) do mesmo, sua base conjugada, que é eliminada na reação, é fraca.

Para avaliar se a metodologia proposta era viável, reações de mioglobina com ácido adípico ativado com EDC/HOBt foram realizadas e comparadas com experimentos de controle. A mioglobina de coração de cavalo utilizada nas reações possui 24 resíduos reativos frente ao ácido adípico ativado, sendo 19 lisinas e 5 serinas. As reações foram efetuadas a 25 °C, e a ativação dos grupos ácidos foi conduzida em frasco separado por 30 minutos antes da proteína ser adicionada, conforme descrito anteriormente. Os excessos molares utilizados foram de 500 vezes de ácido adípico em relação à proteína (correspondente a 20 vezes de excesso para cada resíduo reativo da mioglobina), de 1000 vezes de EDC em relação à proteína (2 vezes em relação ao ácido adípico) e de 4000 vezes de HOBt em relação à proteína (8 vezes em relação ao ácido adípico e 4 vezes em relação ao EDC). Essas condições foram usadas por serem próximas das utilizadas na metodologia utilizando a 1,6-hexanodiamina, e foram posteriormente otimizadas no decorrer do projeto. O maior excesso de HOBt é necessário pois se a espécie ativada somente com EDC sofrer hidrólise no meio reacional, o EDC é perdido na forma de um derivado de ureia, logo deseja-se criar uma condição na qual o HOBt desloque o EDC na ativação o mais rápido possível. No caso do HOBt, ao sofrer a hidrólise esse composto é regenerado, podendo seguir deslocando outras espécies de EDC formadas na primeira etapa da ativação. Os espectros de MS resultantes dessa primeira reação e dos experimentos controle são apresentados na Figura 16.



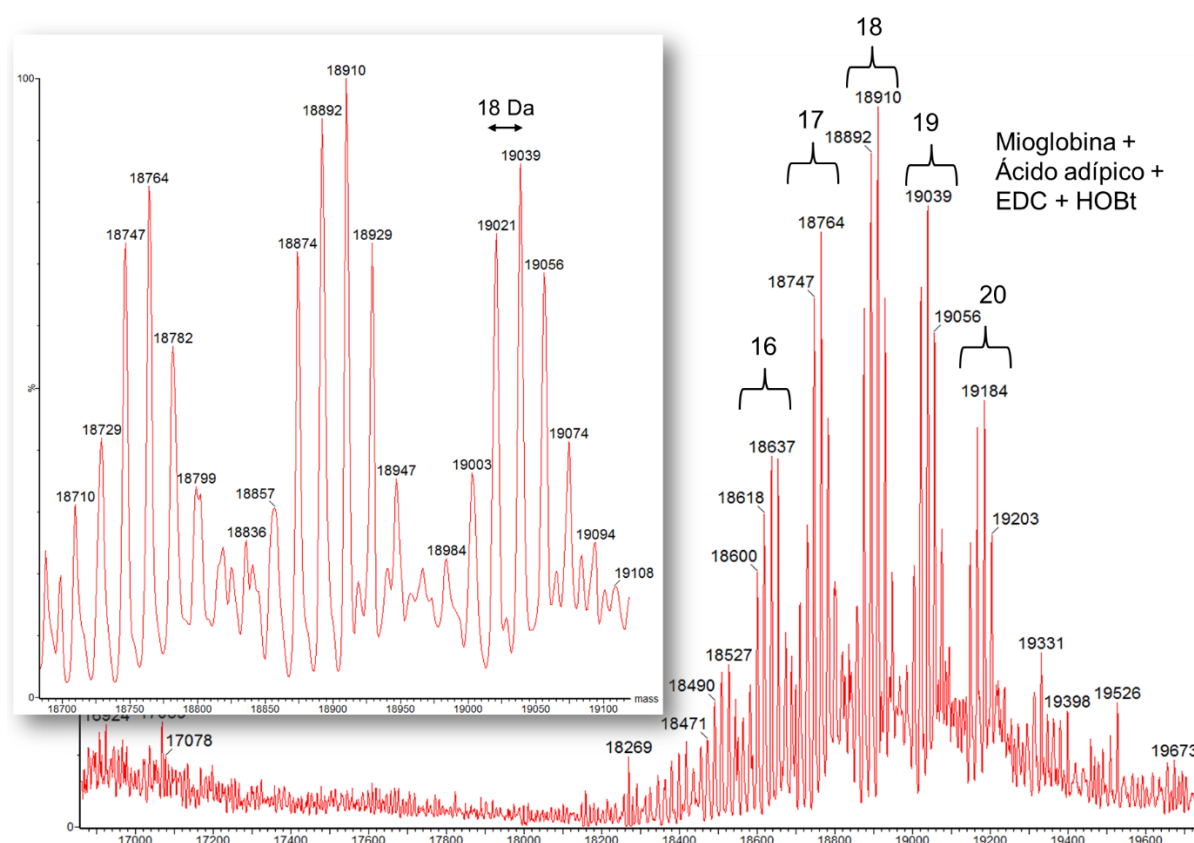
**Figura 16.** Espectros deconvoluídos de ESI(+)-MS de mioglobina intacta na presença e ausência dos ativadores do ácido adípico. Nota-se que a dupla ativação do ácido adípico é realmente necessária para sua incorporação na proteína.

Observa-se que, somente com a adição de ácido adípico à proteína, não houve a formação de nenhuma espécie de ligação cruzada, cuja diferença de massa para esse ALC seria de 110 Da por ALC adicionado à proteína não modificada ( $m/z$  16952). Também não houve a formação de nenhuma espécie do tipo *dead-end*, cuja diferença de massa em relação à proteína não modificada seria de 128 Da. Por fim, também é possível notar que a mioglobina praticamente não foi consumida na reação. Ao se adicionar o EDC, os grupos ácidos foram ativados e ocorreu a incorporação de até duas moléculas de ácido adípico, formando majoritariamente duas espécies *dead-end*, de  $m/z$  17080 e 17207. Pode-se notar, ainda, que o HOBt sozinho não foi capaz



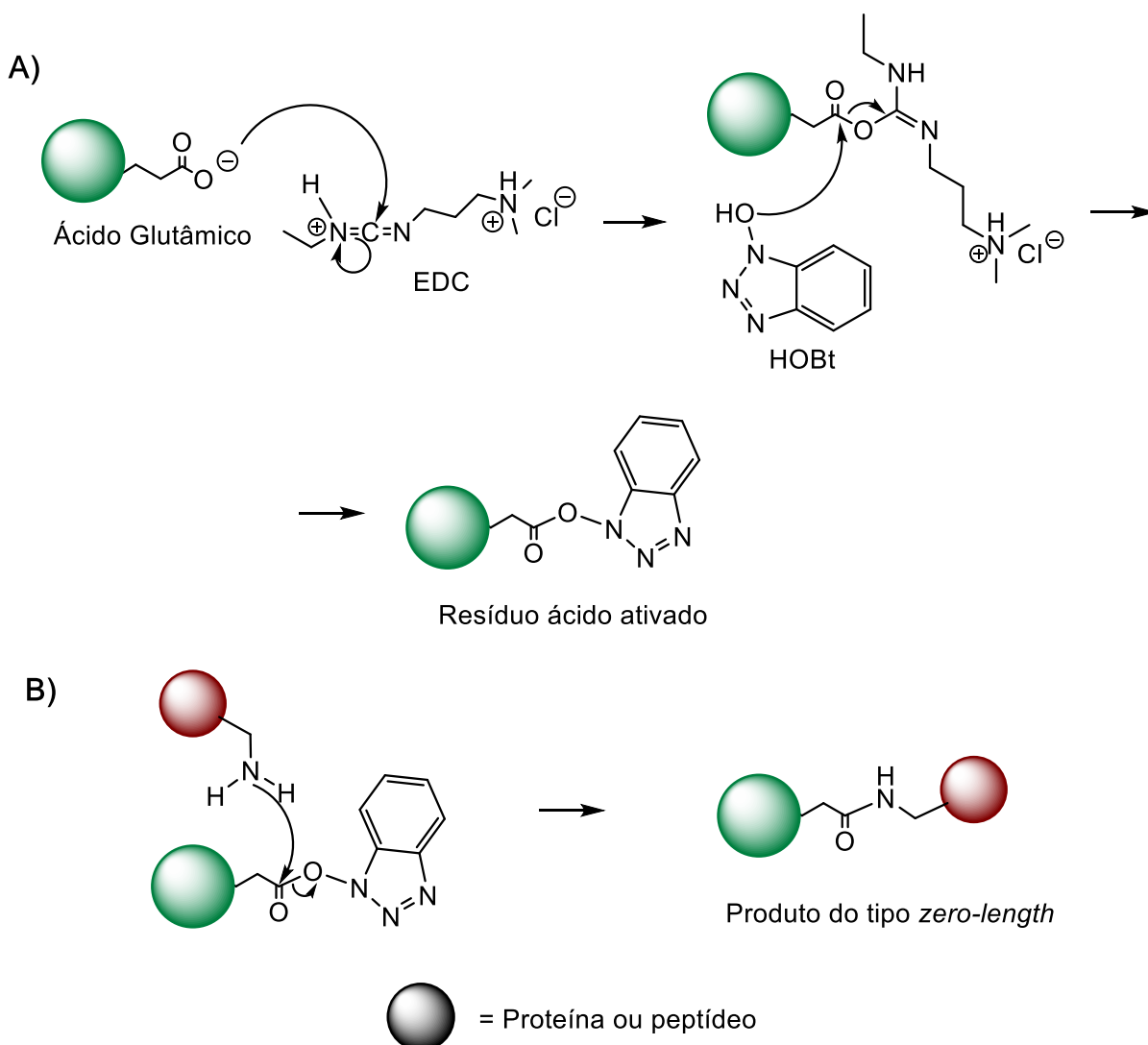
de ativar os grupos ácidos, não havendo incorporações de ácido adípico nem consumo da proteína quando o experimento foi realizado sem a adição de EDC.

Por fim, analisando-se o último espectro da Figura 16, é fácil perceber que a reação de ligação cruzada com a dupla ativação EDC/HOBt ocorreu com bons rendimentos. Nesse cenário, houve a incorporação de até 18 moléculas de ácido adípico, evidenciada pelo sinal de  $m/z$  19203, que pode ser uma combinação de 15 *dead-ends* e 3 ligações cruzadas, por exemplo. Ao ampliar os conjuntos de sinais desse último espectro (Figura 17), é interessante notar a formação de uma série de espécies com diferenças de massa de -18 Da a partir do *dead-end* correspondente, que podem vir da formação de um produto de ligação cruzada diferente dos apresentados até o momento, entre resíduos ácidos e nucleofílicos, do tipo *zero-length*.



**Figura 17.** Ampliação do espectro deconvoluído de ESI(+)-MS da reação de mioglobina com ácido adípico duplamente ativado. Para cada incorporação de ALC, observa-se sinais correspondentes a diferenças de massa de 18 Da, indicando a formação de espécies de ligação cruzada do tipo *zero-length*. No quadro interno, uma ampliação ainda maior mostrando a massa correspondente desses sinais.

O mecanismo de formação das espécies do tipo *zero-length* (Figura 18) é muito semelhante ao do ácido adípico, mas nesse cenário são as cadeias laterais dos resíduos de ácido aspártico e glutâmico da proteína que são ativadas por EDC e HOBt, sendo posteriormente atacadas por resíduos nucleofílicos. O nome dessas espécies vem da ausência de uma cadeia espaçadora que une os dois resíduos reativos, de modo que a restrição de distância entre eles obtida através dessas espécies corresponde apenas ao raio de Van der Waals dos átomos que participam da reação.

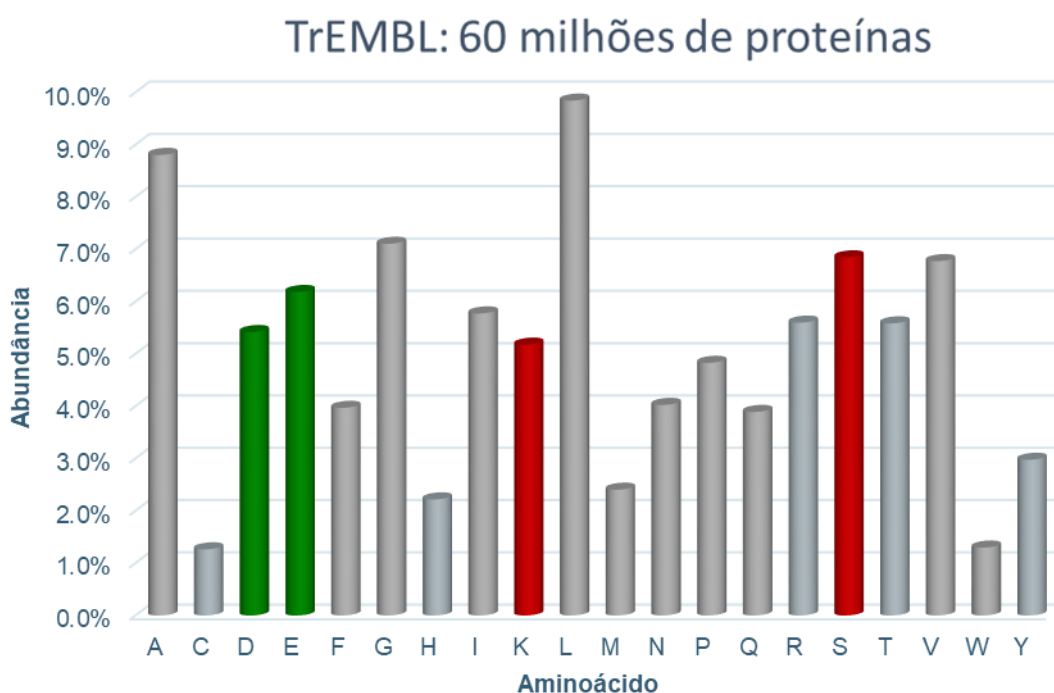


**Figura 18.** Mecanismo de reação para formação de produtos de ligação cruzada do tipo *zero-length*. A) ativação dos resíduos ácidos (representado pelo ácido glutâmico) com EDC e HOBt; B) reação do resíduo ácido ativado com resíduo nucleofílico próximo (representado por lisina) para formação do produto de ligação cruzada.

Com a possibilidade de formação conjunta das espécies do tipo *zero-length* (ZL) e de ligação cruzada inter ou intramoleculares (XL), analisando-se novamente o

último espectro da Figura 16, pode-se perceber que existe um grande número de combinações possíveis de espécies que resultariam nos sinais obtidos. O sinal de  $m/z$  19203, que havia sido atribuído à presença de 15 *dead-ends* e 3 XL, agora pode ser também fruto de uma combinação de 16 *dead-ends*, 2 XL e 1 ZL, 17 *dead-ends*, 1 XL e 2 ZL, entre outras. Os sinais de  $m/z$  mais baixos, com 15, 16 e 17 incorporações ( $m/z$  entre 17500 e 18000), possuem uma variedade ainda maior de combinações possíveis, com maior número de XL formadas em detrimento de *dead-ends*, maior número de ZL, e assim por diante. A atribuição exata de quais espécies estão sendo formadas na reação de ligação cruzada se dá a nível de MS/MS (ou MS<sup>2</sup>), mas os espectros de MS fornecem informações a respeito do número de incorporação média do ALC por molécula de proteína, sendo úteis para avaliar a eficiência da metodologia desenvolvida em diferentes condições, que serão discutidas adiante.

A possibilidade de obtenção de restrições de distância agora envolvendo também os resíduos ácidos constitui uma vantagem muito importante da metodologia desenvolvida quando comparada com os ésteres de NHS devido ao maior número de combinações possíveis entre resíduos para formação das espécies de ligação cruzada. A Figura 19 mostra a abundância relativa dos resíduos de aminoácidos, de acordo com estatística do banco de dados TrEMBL<sup>62</sup>, contendo aproximadamente 60 milhões de proteínas.

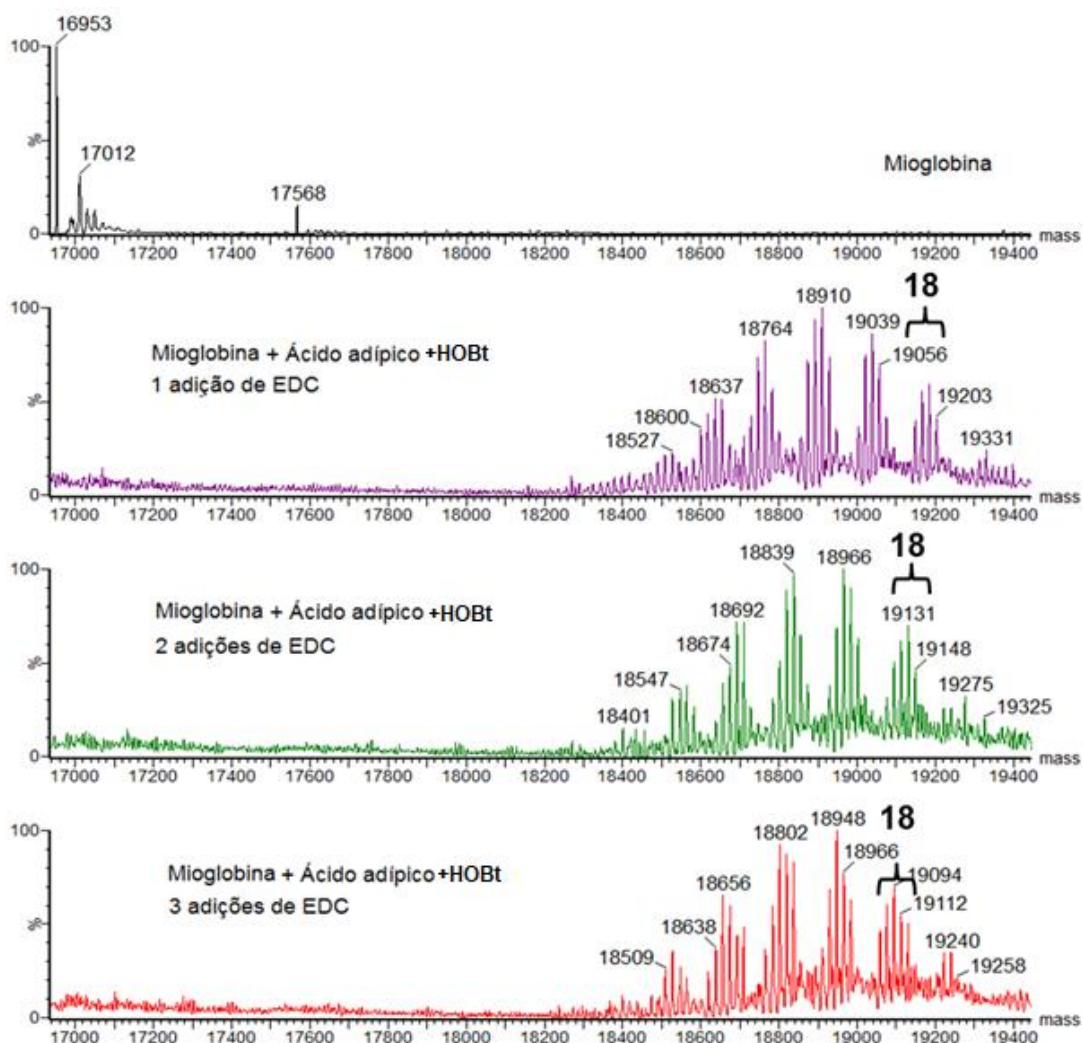


**Figura 19.** Gráfico apresentando a abundância relativa dos resíduos de aminoácidos no conjunto de aproximadamente 60 milhões de proteínas do banco de dados TrEMBL.

Na metodologia dos derivados de NHS os resíduos reativos são lisinas (K) e serinas (S) que, somados, correspondem a aproximadamente 12% dos aminoácidos presentes nas proteínas. Além disso, 3 tipos de espécies de ligação cruzada podem ser obtidos de acordo com os resíduos reativos: K-K; K-S e S-S. Ao possibilitar a formação de espécies *zero-length*, a metodologia do ácido adípico inclui também os resíduos de ácido aspártico (D) e glutâmico (E) como resíduos reativos, sendo que estes somados possuem abundância de pouco mais de 11%. Assim, a frequência de resíduos reativos nesta metodologia passa de 12 para mais de 23%, praticamente dobrando a abundância de resíduos amostrados de uma proteína. Somando-se a isso, além dos mesmos 3 tipos de espécies formadas pela reação do ácido adípico, na reação de *zero-length* há a possibilidade de formação de mais 4 espécies: K-D; K-E; S-D e S-E. Com quase o dobro da abundância dos resíduos amostrados e possibilidade de formação de mais do dobro de tipos de espécies, essa metodologia tem capacidade de fornecer um conjunto muito maior de informações estruturais a respeito da proteína estudada.

#### **4.3. Otimização da metodologia de ligação cruzada utilizando o ácido adípico**

Uma vez que foi possível incorporar moléculas do ácido adípico na mioglobina e que para isso foi necessária uma dupla ativação dos grupos ácidos com EDC e HOBt, foram realizados vários testes com o objetivo de otimizar as condições reacionais. O primeiro foi efetuar sucessivas adições de EDC no decorrer da reação. Esse procedimento foi feito pois o aduto de EDC por ser hidrolisado no meio reacional e consumido (adições extras de HOBt não são necessárias pois ele é regenerado em caso de hidrólise de seu aduto) durante o período de reação, que normalmente é de duas horas. Neste teste foi analisada a influência de 3 adições de EDC, em intervalos de 30 minutos, totalizando uma hora e meia de reação (Figura 20). Essa reação foi efetuada com excesso molar também de 500 vezes de ácido adípico em relação à proteína, mantendo as mesmas proporções de EDC e HOBt em relação ao diácido.



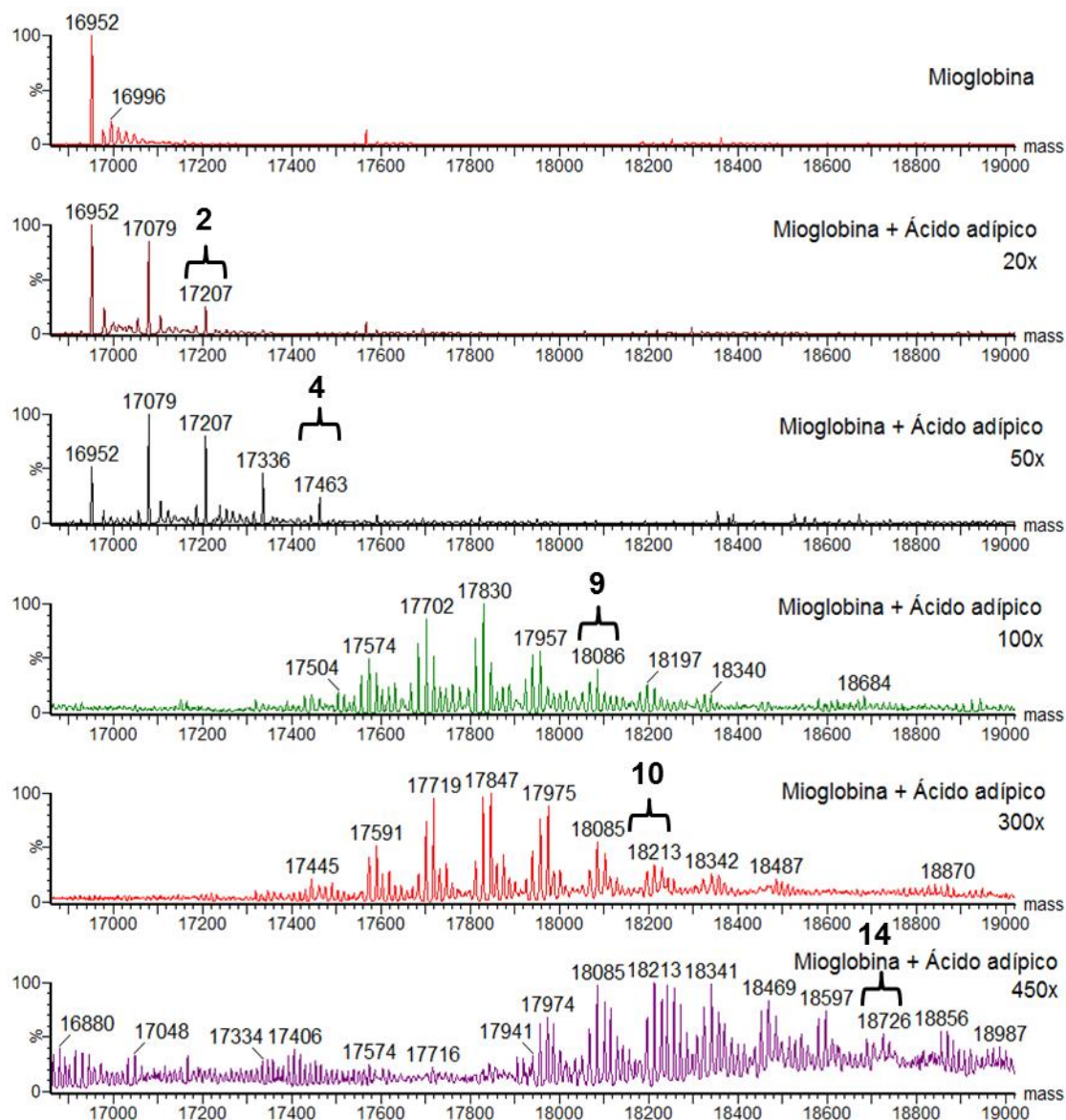
**Figura 20.** Espectros deconvoluídos de ESI(+)-MS dos produtos de ligação cruzada com várias adições de EDC a cada 30 minutos. Observa-se que não há aumento do rendimento de formação de ligação cruzada com adições extras de EDC ao meio reacional.

Como pode-se observar, o perfil de incorporação do ácido adípico na mioglobina foi o mesmo para as três adições de EDC, indicando que o ataque do HOBt é mais rápido que a taxa de hidrólise do aduto de EDC. Nos três casos, houve a adição de até 18 moléculas do ALC, mesmo resultado visto no último espectro da Figura 16. Desse modo, para as reações seguintes, foi adotada como padrão uma única adição de EDC, bem como a diminuição do tempo reacional para apenas 30 minutos. A redução do tempo de reação possui vantagens associadas a qualidade da informação estrutural que é possível se obter do sistema de interesse, uma vez que há diminuição na probabilidade de uma proteína-alvo sofrer desenovelamento em solução, além de permitir que a técnica de ligação cruzada seja aplicada a proteínas que não seriam

estáveis períodos longos, mas que podem ser submetidas aos experimentos em intervalos de tempo menores.

Estando estabelecido um menor tempo reacional, o teste seguinte consistiu em testar a influência do excesso molar do ALC em relação à proteína na taxa de incorporação, visando a redução da quantidade de reagentes utilizada e também de incorporações, já que estudos por IM-MS realizados em nosso grupo de pesquisa mostraram que incorporações acima de 10 moléculas do ALC podem causar alterações na seção de choque da mioglobina, indicando mudanças conformacionais mais severas em sua estrutura (dados não publicados). Ao reduzir o número de incorporações aumenta-se a confiabilidade dos dados de distância obtidos, que passam a refletir com maior fidelidade a estrutura da proteína em solução. No entanto, existe um compromisso entre a taxa de incorporação e a detecção dos peptídeos modificados, de modo que se a taxa de incorporação do ALC for muito baixa, os sinais dessas espécies se tornam muito diluídos frente aos peptídeos não modificados, tornando muito difícil sua detecção e identificação.

Nesses experimentos, foram testados os excessos molares de ácido adípico de 20, 50, 100, 300 e 450 vezes em relação à mioglobina, mantendo-se a proporção 2:1 de EDC e 8:1 de HOBt em relação ao diácido em todos os casos (Figura 21).



**Figura 21.** Espectros deconvoluídos de ESI(+)-MS da reação da mioglobina com diferentes excessos de ácido adípico. Observa-se que a taxa de incorporação média de moléculas de ALC aumenta com os excessos molares de ácido adípico, atingindo valores ideais com 100 e 300 vezes de excesso. Em todas as reações as proporções de EDC e HOBt foram mantidas em 2:1 e 8:1 em relação ao ALC, respectivamente.

Ao se utilizar um excesso molar de 20 vezes do ácido adípico em relação à proteína, houve a formação de apenas duas espécies do tipo *dead-end*, de  $m/z$  17079 e 17207, sem que tenha ocorrido a formação de nenhuma espécie XL ou ZL. Aumentando-se o excesso molar do ácido adípico para 50 vezes, o número de incorporações dobrou, mas as espécies formadas foram apenas do tipo *dead-ends*, ainda sem XL. Ao se elevar o excesso molar do ALC para 100 vezes, no entanto, nota-se um salto considerável no rendimento da reação de ligação cruzada. Além do



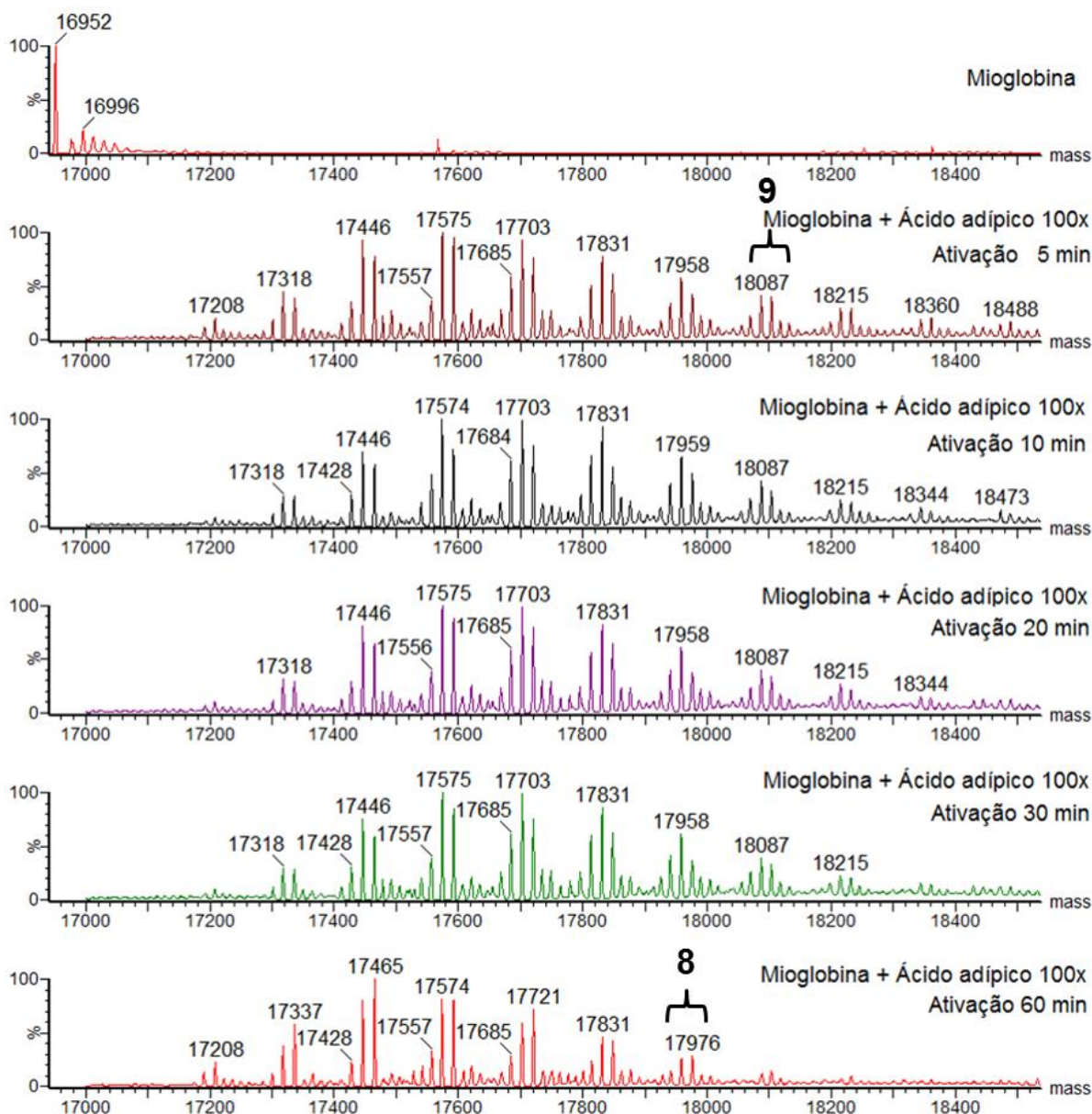
aumento na taxa de incorporações, chegando a até 9 (sinal de  $m/z$  18086, que pode ser correspondente a 8 *dead-ends* e 1 XL, por exemplo), houve a formação de diversas espécies cujos sinais são correspondentes à proteína modificada por múltiplos de 110 Da, indicando que, nessa condição, ocorreram incorporações de ácido adípico nas quais ambos os grupos ácidos estavam ativados e reagiram com dois diferentes resíduos nucleofílicos, formando espécies de ligação cruzada. Também pode-se notar que, ao contrário das condições com 20 e 50 vezes de excesso, com um excesso molar de 100 vezes toda a proteína foi consumida na reação, o que é evidenciado pela ausência do sinal de  $m/z$  16952. Utilizando-se um excesso molar de 300 vezes do diácido não houve uma mudança significativa no perfil de incorporação do ALC quando comparado com o experimento em que foi utilizado o excesso de 100 vezes, com apenas uma incorporação a mais. Já ao se utilizar um excesso molar de 450 vezes de ácido adípico, observa-se uma taxa de incorporação ainda maior, de até 14 moléculas, com formação das diferentes espécies de *dead-end*, XL e ZL.

Para o decorrer do desenvolvimento da metodologia, foi escolhida como a melhor condição para a reação a utilização de excesso molar de ácido adípico de 100 vezes pois, quando comparada com os ensaios com 20 e 50 vezes de excesso, nesta condição há a formação de múltiplas espécies XL e ZL em detrimento de apenas poucas espécies *dead-end* das outras duas condições. Ao contrário dessas espécies hidrolisadas que acabam fornecendo apenas informações sobre a acessibilidade do resíduo que reagiu frente ao solvente, as espécies inter e intra peptídicas fornecem informações acerca da distância máxima na estrutura terciária entre os dois resíduos que reagiram, sendo essa a informação desejada para a etapa de modelagem computacional afim de gerar modelos que representem a estrutura da proteína em solução. Sendo assim, é fundamental que essas espécies estejam sendo formadas nos ensaios de ligação cruzada. Quando comparada com os ensaios utilizando maiores excessos, a condição com 100 vezes foi escolhida por apresentar uma considerável economia de reagentes com mesma incorporação frente a condição de 300 vezes, além da taxa de incorporação do ALC estar dentro dos valores considerados ideais quando comparada com a condição de excesso de 450 vezes (e ainda mais nas condições iniciais de 500 vezes de excesso), condições em que a incorporação foi muito elevada. Embora 100 vezes pareça um excesso grande do ALC, é importante lembrar que cada molécula de mioglobina possui 24 resíduos



nucleofílicos de lisina e serina, de modo que o excesso molar do ácido adípico é, na verdade, de apenas 4 vezes em relação a cada resíduo reativo da proteína.

Outro parâmetro testado no desenvolvimento da metodologia foi a influência do período de ativação dos grupos ácidos antes de se adicionar a mioglobina nas taxas de incorporação do ALC. Esse procedimento de realizar a ativação separadamente foi empregado como uma tentativa de ativar preferencialmente os grupos ácidos do ALC em detrimento dos resíduos ácidos da proteína, favorecendo assim a obtenção de espécies XL. Para realização do teste, uma solução contendo EDC, HOBt e ácido adípico dissolvidos no mesmo tampão MOPS de reação foi preparada e a ativação procedeu pelos períodos de 5, 10, 20, 30 e 60 minutos, antes que a mioglobina fosse adicionada. Os resultados são apresentados na Figura 22.



**Figura 22.** Espectros deconvoluídos de ESI(+)-MS da reação da mioglobina utilizando diferentes períodos de ativação do ácido adípico. Observa-se que o padrão de incorporação do ALC não se altera significativamente em nenhuma das condições, apresentando uma incorporação em média com ativação de 60 minutos.

Observando-se os espectros, é possível notar que não houve diferença significativa na taxa de incorporação do ácido adípico na mioglobina em nenhum dos períodos avaliados. A diferença mais perceptível ocorreu com o período de ativação mais longo, de 60 minutos, no qual foi observada, em média, uma incorporação a menos do ALC. O motivo dessa redução provavelmente decorre do fato de o EDC ser perdido no meio reacional caso seu aduto sofra hidrólise. Ao permitir que essas espécies permanecessem em solução por mais tempo, a probabilidade de uma das

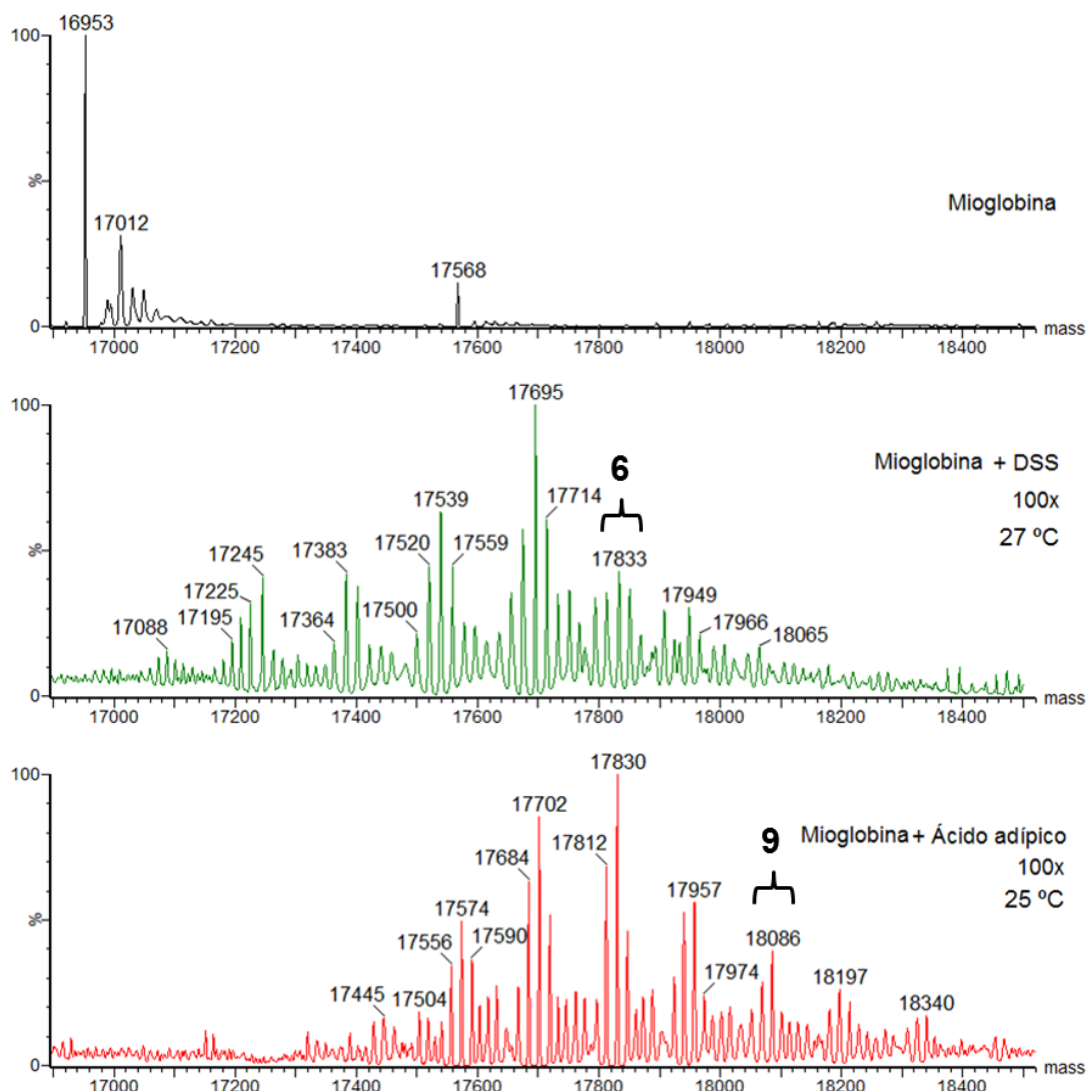
extremidades ter sido hidrolisada foi maior, restando menos espécies do aduto de EDC para serem atacadas pelo HOBt e, conseqüentemente, menos grupos ativados para reagirem com os resíduos da mioglobina.

Como não houve diferença nas taxas de incorporação, para todas as reações seguintes a este teste foi adotado um período de ativação de 5 minutos. Assim, após a realização de todos esses testes, as condições estabelecidas para os experimentos de ligação cruzada utilizando o ácido adípico como ALC envolveram um excesso molar de 100 vezes do diácido em relação à proteína e uma proporção de EDC e HOBt de 2:1 e 8:1, respectivamente, em relação ao diácido. A ativação dos grupos ácidos foi conduzida em um frasco separado não contendo a proteína, por 5 minutos, e a reação após a adição da mesma foi realizada em 30 minutos.

#### **4.4. Estudo comparativo entre os experimentos de ligação cruzada utilizando o ácido adípico frente o uso de DSS**

Uma vez que a metodologia foi otimizada e apresentou resultados bastante satisfatórios, foram realizados ensaios comparativos tendo como base o uso de DSS como ALC, que é atualmente o reagente padrão utilizado em XL-MS.

O primeiro experimento foi realizado nas condições padrões da reação com DSS, isto é, um excesso molar de 100 vezes em relação à proteína, a 27 °C, por duas horas. Conforme mencionado anteriormente, o DSS foi solubilizado em DMF e posteriormente adicionado à solução de mioglobina, de modo que o volume de DMF não passasse de 5% do volume total da reação. O espectro obtido, juntamente ao espectro obtido utilizando-se o ácido adípico nas condições otimizadas e a 25 °C para efeito de comparação são apresentados na Figura 23.

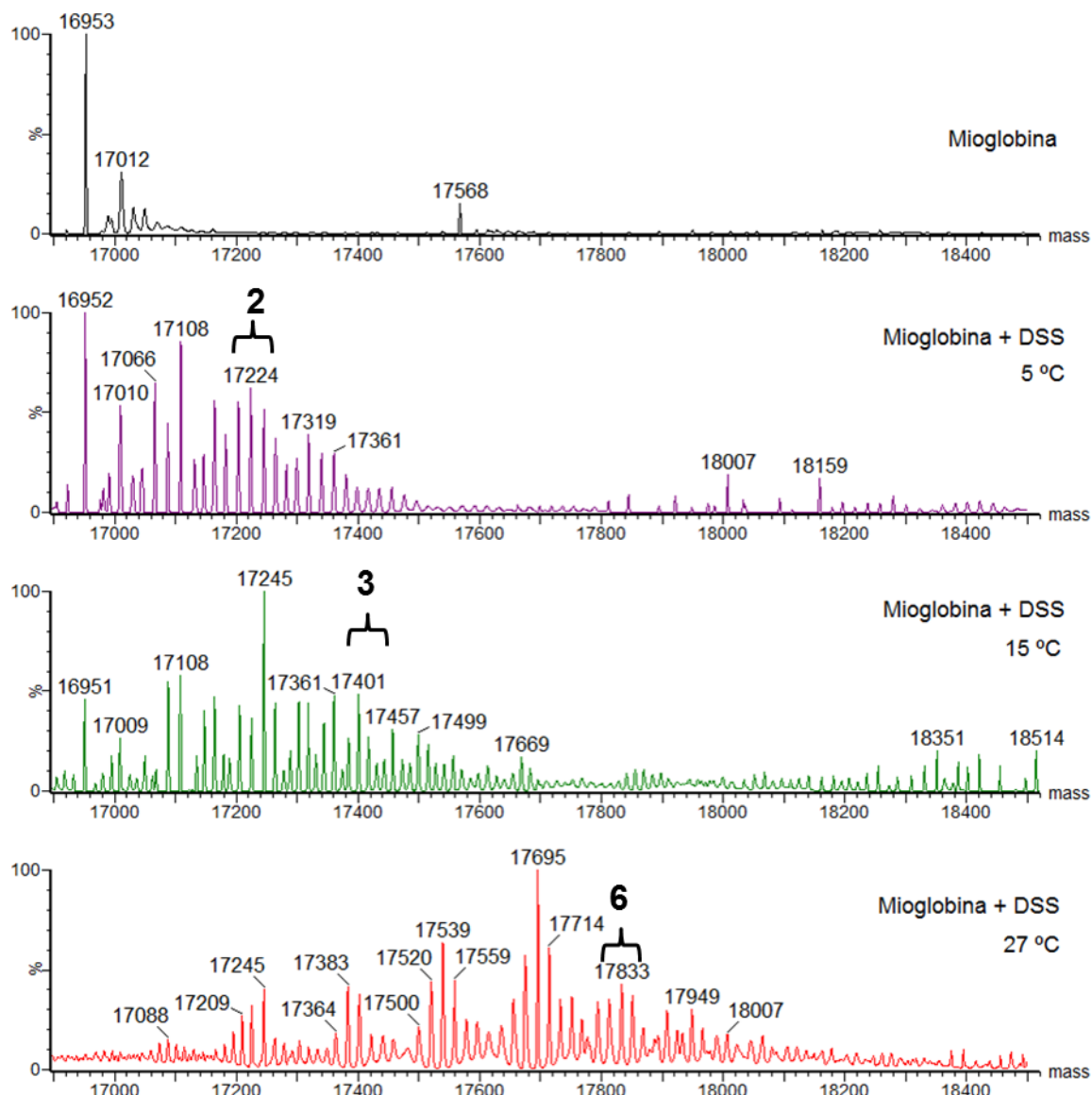


**Figura 23.** Espectros deconvoluídos de ESI(+)-MS de mioglobina comparando as reações de ligação cruzada utilizando o DSS e o ácido adípico como ALC. Nota-se que, mesmo na temperatura ótima da reação com DSS, a taxa de incorporação do ALC por molécula da proteína é consideravelmente menor que a do ácido adípico.

Observando-se a figura, nota-se que a incorporação média para a reação com DSS foi de 6 moléculas, correspondente ao sinal de  $m/z$  17833 que pode ser atribuído a 4 espécies *dead-end* e 2 espécies XL (para o DSS as adições de massa para *dead-end* e XL são de 138 e 156 Da, respectivamente), enquanto que a reação com o ácido adípico teve uma taxa de incorporação média de 9 moléculas do ALC. Comparando-se as duas reações, o uso do ácido adípico como uma alternativa ao DSS parece realmente promissor. Com efeito, o fato do ácido adípico ter mostrado maior rendimento chega a ser surpreendente, uma vez que o DSS possui uma cadeia espaçadora maior (11,7 Å do DSS contra 9,1 Å do ácido adípico), ou seja, a faixa de

distâncias entre resíduos permitidas para que o DSS fosse capaz de uni-los é maior. Utilizando-se um software desenvolvido por nosso grupo de pesquisa em parceria com o Dr. Leandro Martínez (IQ/UNICAMP-SP), chamado Topolink<sup>63</sup>, que calcula a distância topológica (distância física) entre os resíduos reativos de acordo com a especificidade e tamanho do ALC dada a estrutura da proteína, foram calculadas 62 espécies XL para o DSS e 44 para o ácido adípico, este também tendo a possibilidade de formar até 54 espécies ZL. Esse fato evidencia ainda mais uma maior eficiência da metodologia proposta.

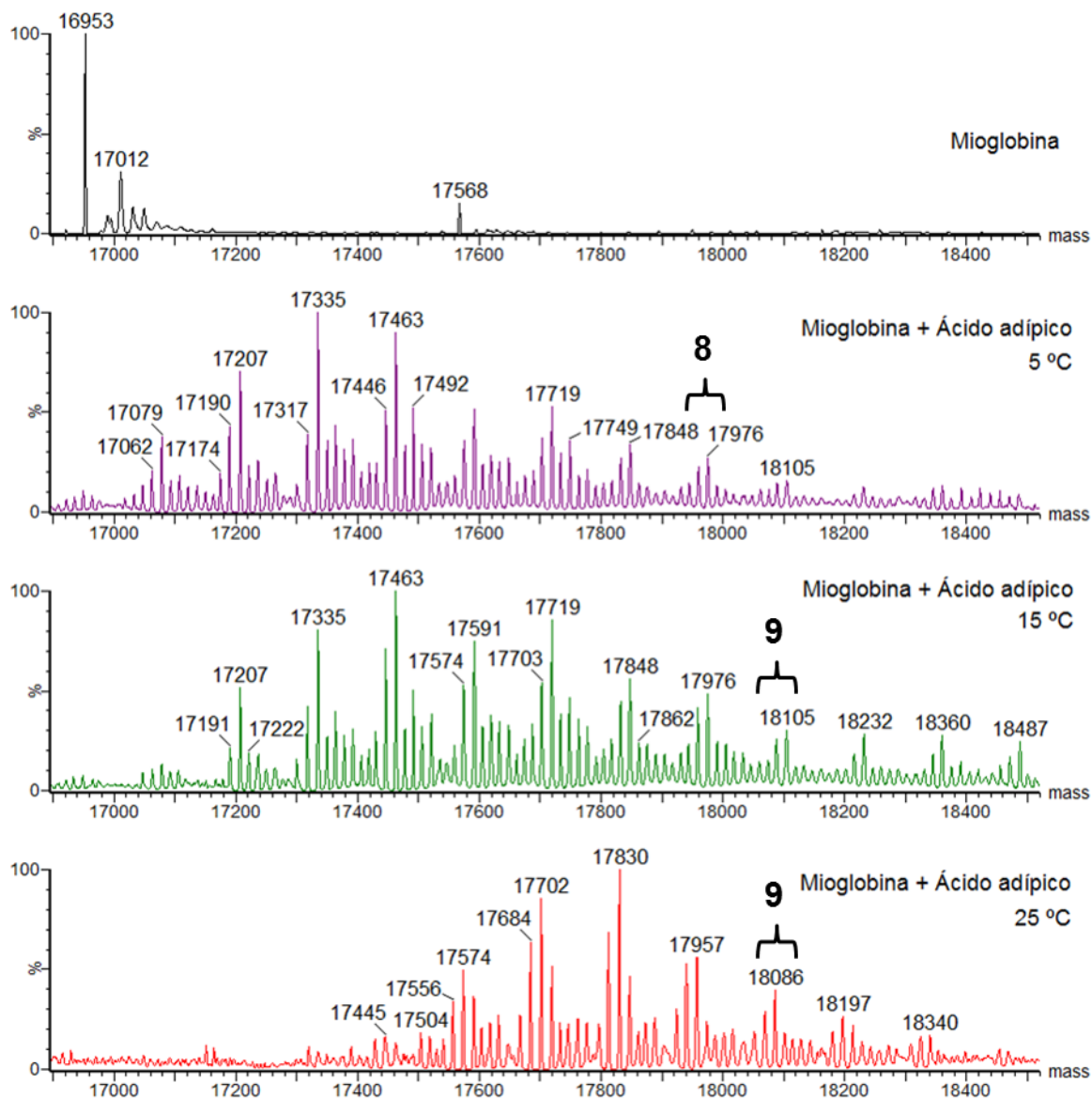
Uma das grandes limitações do uso de DSS (e dos derivados de ésteres de NHS em geral) como ALC, como já mencionado, é a queda de rendimento da reação de ligação cruzada em temperaturas mais baixas. Essa limitação inviabiliza sua aplicação em proteínas sensíveis à temperatura, que podem sofrer desenovelamento/agregação com o aumento da mesma, causando uma redução na aplicabilidade da técnica de XL-MS como um todo. Para confirmar esta limitação do DSS frente à redução de temperatura, foram realizados ensaios de ligação cruzada com mioglobina a 15 e a 5 °C. Os espectros são apresentados na Figura 24, e o espectro da reação a 27 °C também foi adicionada para efeitos de comparação.



**Figura 24.** Espectros deconvoluídos de ESI(+)-MS de mioglobina nas reações de ligação cruzada para avaliação da influência da temperatura na taxa de incorporação de DSS. Observa-se que, com o decréscimo da temperatura, a incorporação do ALC cai drasticamente.

Como é facilmente observado, o rendimento da reação de ligação cruzada com o DSS foi muito afetado pela redução da temperatura. Se a taxa de incorporação média foi de 6 moléculas do ALC a 27 °C, reduzindo-se a temperatura para 15 °C somente três incorporações ocorreram em média, dadas pelo sinal de  $m/z$  17401, correspondente a 2 *dead-ends* e 1 XL. Já a 5 °C o rendimento caiu ainda mais, com a formação apenas de 1 *dead-end* e 1 XL ( $m/z$  17246, sinal imediatamente a direta do  $m/z$  17224). Logo, a utilização do DSS como agente de ligação cruzada para estudos de proteínas sensíveis à temperatura se mostra extremamente inviável.

As mesmas temperaturas de reação foram testadas, agora utilizando o ácido adípico como ALC, e os espectros são apresentados na Figura 25, além do espectro da reação a 25 °C.



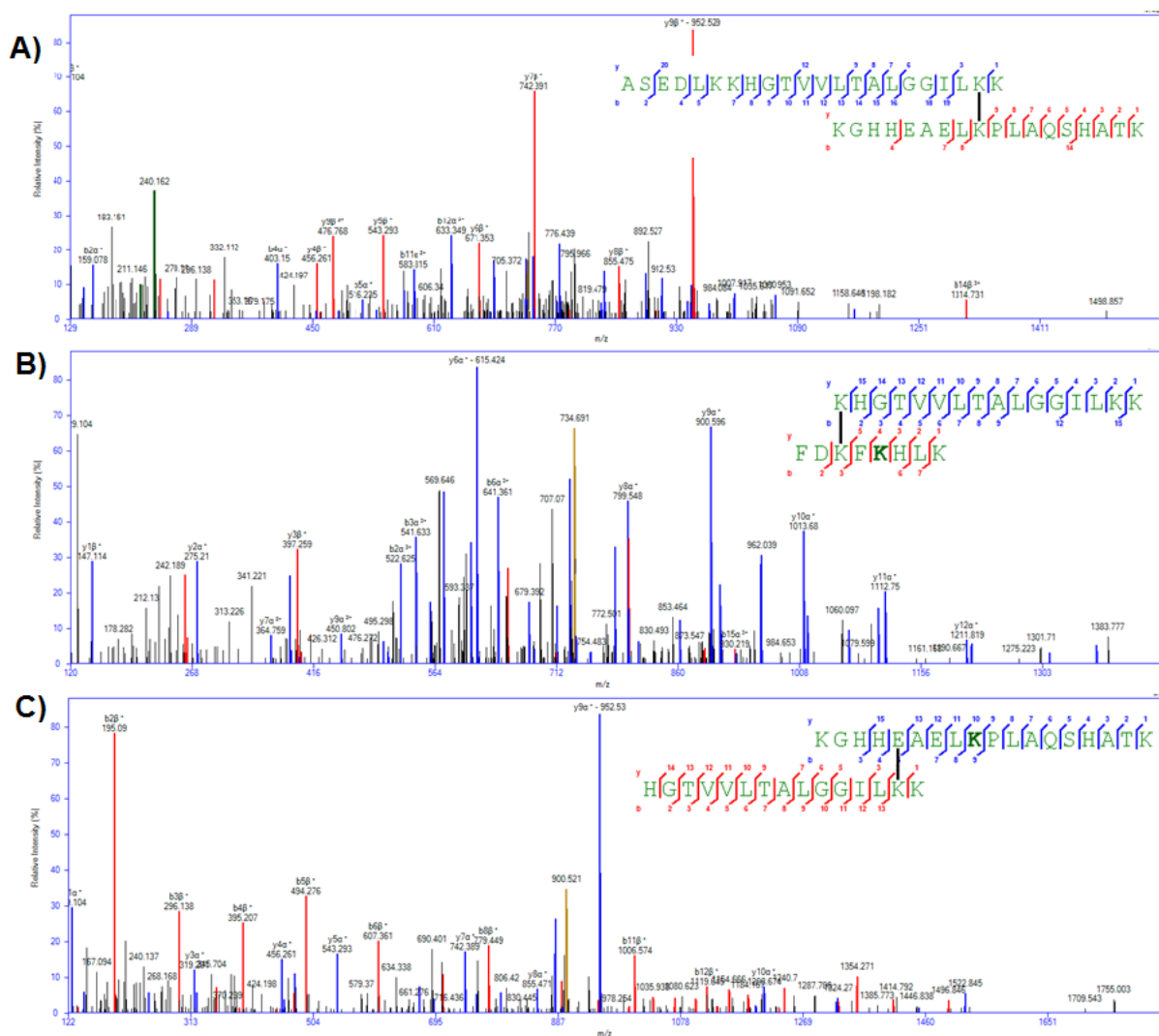
**Figura 25.** Espectros deconvoluídos de ESI(+)-MS de mioglobina nas reações de ligação cruzada para avaliação da influência da temperatura na taxa de incorporação do ácido adípico. Observa-se que, mesmo com o decréscimo da temperatura, a taxa de incorporação do ALC se mantém praticamente inalterada.

A diferença entre as reações com DSS e ácido adípico no que diz respeito à sensibilidade frente a redução de temperatura foi muito grande. A 15 °C, foi observada a mesma taxa de incorporação de ácido adípico que na reação conduzida a 25 °C, a diferença estando apenas na presença também de espécies com 2, 3 e 4 incorporações ( $m/z$  17207, 17335 e 17463, sinais que correspondem somente a

espécies do tipo *dead-ends*, além dos sinais das possíveis combinações de espécies XL e ZL, que diferem desses  $m/z$  em múltiplos de 18 Da). Mesmo ao reduzir a temperatura reacional para 5 °C, foi observada ainda uma taxa de incorporação de até 8 ALCs, evidenciada pelo sinal de  $m/z$  17976 (que corresponde a 8 *dead-ends*, além das espécies de XL e ZL adjacentes). Esses resultados mostram que, para a mioglobina, mesmo a 5 °C – uma temperatura já bem próxima de zero – a taxa de incorporação do ácido adípico é maior do que a do DSS em suas condições ótimas, à 27 °C.

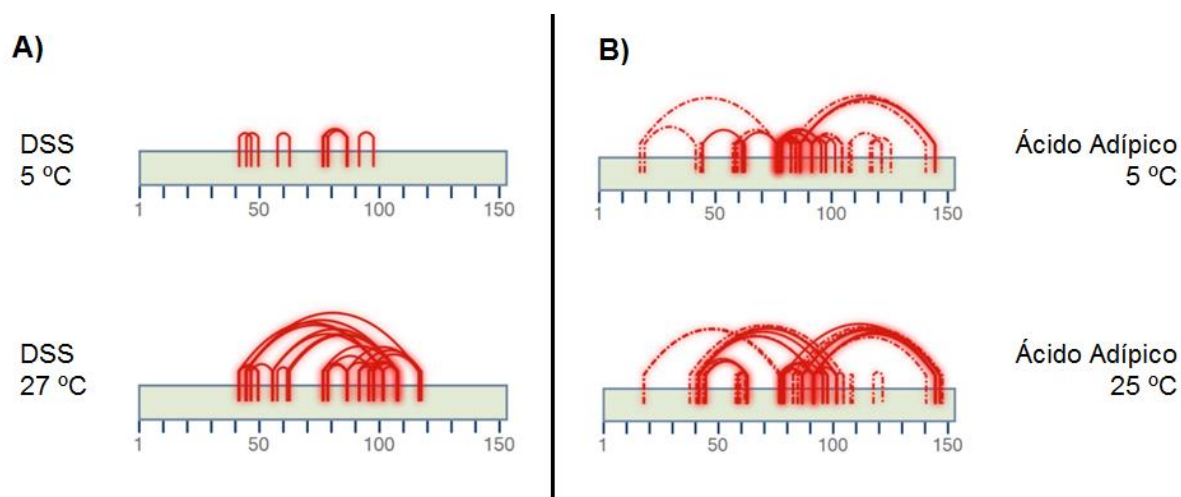
Apesar dos espectros de MS fornecerem as informações a respeito do tipo das espécies formadas e da taxa de incorporação média do ALC por proteína (que pode ser interpretada como uma forma de avaliar a eficiência da reação de ligação cruzada), a identificação do número de espécies formadas, assim como sua identidade (e, portanto, as restrições de distância entre os resíduos reativos) são obtidos a nível de MS<sup>2</sup>. No caso da mioglobina, que possui estrutura muito bem conhecida, as restrições de distância obtidas não foram utilizadas para guiar uma etapa de modelagem computacional, mas esse seria o caso se uma proteína de estrutura desconhecida fosse o alvo. A Figura 26A mostra um espectro de íons produto, obtido em experimentos de MS<sup>2</sup>, de uma espécie de ligação cruzada intermolecular para a reação com DSS, enquanto que as Figura 26B e Figura 26C mostram espectros de íons fragmento de espécies de ligação cruzada intermolecular do tipo XL e ZL, respectivamente, para a reação com ácido adípico.





**Figura 26.** Exmplos de espectros de íons fragmentos anotados para espécies de ligação cruzada. A) espécies XL entre lisinas através de reação com DSS; B) espécies XL entre lisinas através de reação com ácido adípico; C) espécies ZL entre lisina e ácido glutâmico através de reação com ácido adípico.

A distribuição de todas as restrições obtidas através das espécies de ligação cruzada pela sequência da proteína pode ser visualizada através de um mapa de restrições (Figura 27).



**Figura 27.** Mapa das ligações cruzadas identificadas para a mioglobina à 25 e 5 °C com DSS (A) e ácido adípico (B). Nota-se a formação de mais espécies nas duas reações em que foi utilizado o ácido adípico como ALC, mesmo em temperatura reduzida. Também é possível notar uma cobertura maior da sequência da proteína.

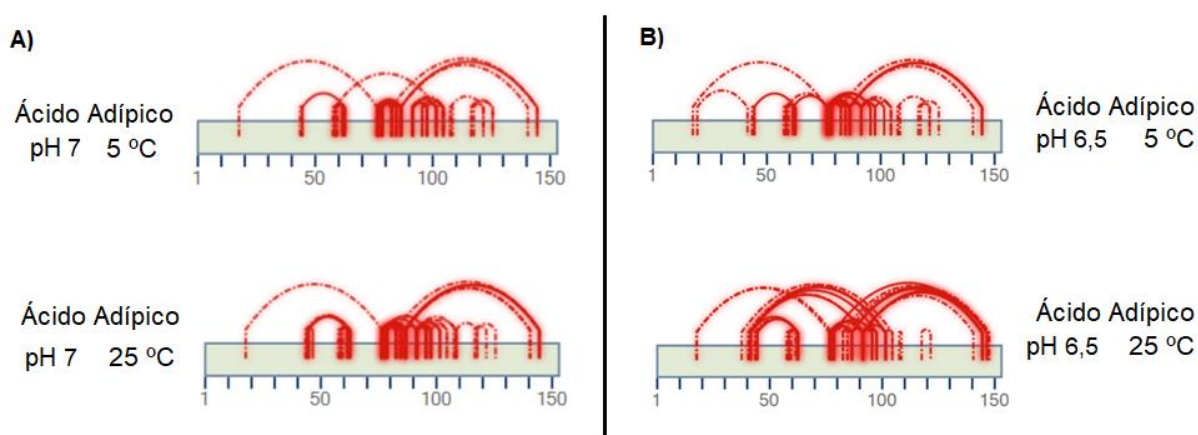
É fácil notar que as reações envolvendo o DSS como ALC realmente são muito ineficientes em temperaturas reduzidas, sendo que a 5 °C foram formadas apenas 6 espécies XL, contra 19 espécies a 27 °C. No caso do ácido adípico, foram 49 espécies formadas a 25 °C, sendo 25 XL e 24 ZL, e a reação continuou com boa taxa de formação mesmo a 5 °C, com 18 XL e 24 ZL, totalizando 42 espécies de ligação cruzada. Além disso, considerando os experimentos a 25 e 27 °C, a cobertura da sequência da proteína foi maior na reação do ácido adípico, e uma amostragem mais abrangente da sequência auxilia a geração de melhores modelos na etapa de modelagem. Esses resultados novamente apontam para o fato de que a metodologia proposta nessa dissertação pode ser aplicada com bons rendimentos mesmo em baixas temperaturas, podendo expandir a aplicabilidade de XL-MS para proteínas instáveis em temperaturas mais elevadas, que até então não eram passíveis de serem estudadas com a utilização de DSS.

Um dado interessante obtido ao se comparar as espécies formadas nas duas reações do ácido adípico é que dentre as 7 espécies do tipo XL que deixaram de ser formadas na reação a 5 °C, 5 delas tinham serinas como um dos resíduos reativos (Tabela 1). A queda no número de espécies do tipo XL formadas é esperada, pois com a diminuição da temperatura o número de colisões efetivas do nucleófilo com as espécies ativadas é menor, uma vez que a distribuição de velocidades de Maxwell-Boltzmann das moléculas é deslocada para médias menores<sup>64</sup>. Dessa forma, é mais

difícil que a barreira de ativação da reação seja vencida, e os ataques nucleofílicos são desfavorecidos. Como a serina é um pior nucleófilo, a barreira de ativação de sua reação é maior, logo com a redução da temperatura e das colisões efetivas um menor número de espécies tende a ser formado. No caso da lisina, a redução das colisões efetivas também deve resultar em um menor número de espécies, mas como ela é um melhor nucleófilo, a barreira de ativação da reação é menor e, portanto, mais facilmente vencida. Logo, a redução no número de espécies formadas não é tão grande como no caso das serinas, e a redução da temperatura acaba tornando a reação mais seletiva, de modo que a reação do pior nucleófilo é mais prejudicada em temperaturas menores e a formação de seus produtos é menor. A manutenção do número de espécies do tipo *zero-length* formadas com a diminuição da temperatura pode ser atribuída ao fato de que, mesmo com a redução das colisões efetivas, a formação dessas espécies é entropicamente muito favorável, já que os dois resíduos estão em contato na superfície da proteína formando pontes salinas (entre ácido e lisina) ou ligações de hidrogênio (entre ácido e serina). Uma vez que o resíduo ácido tenha sido ativado, o número de choques efetivos que ocorrem entre esses dois resíduos é muito maior e, conseqüentemente, o ataque por parte do resíduo nucleofílico tem maior probabilidade de ocorrer, formando o produto.

Em virtude desses dados mostrando que a redução da temperatura desfavorece a formação das espécies do tipo XL e não tem efeito sobre as espécies do tipo ZL, foi realizado um experimento visando avaliar se a variação de pH do meio reacional também teria alguma influência nas espécies formadas. Como pode-se notar pelos mecanismos das reações (Figura 15 e Figura 18), a etapa de ativação é favorecida em meio ácido, já que o ataque do carboxilato ao EDC é facilitado quando este se encontra na forma protonada. Devido à essa etapa de ativação, as reações anteriores a este teste foram todas realizadas em pH=6,5. No entanto, ainda de acordo com os mecanismos apresentados, o ataque da lisina às espécies ácidas ativadas só ocorre em sua forma desprotonada, pois somente nela o nitrogênio possui o par de elétrons disponível para o ataque. Desse modo, um aumento do pH na etapa de reação seria interessante, pois mais espécies de lisina estariam desprotonadas e o ataque nucleofílico seria favorecido (em serinas, nas condições de reação cruzada, a cadeia lateral está sempre na forma R-OH, de modo que a variação do pH não deve ter um impacto grande para esses resíduos). Como uma forma de avaliar a influência do pH nessa segunda etapa, as ativações foram mantidas em pH=6,5, mas a reação

após a adição da mioglobina foi realizada em pH=7. Seria possível avaliar uma variação maior de pH entre as reações, mas procurou-se não variar drasticamente o pH da reação para que a mudança no meio não ocasionasse mudanças conformacionais significativas na proteína a ponto de reduzir a confiabilidade dos dados e as duas metodologias pudessem ser comparadas. Talvez para a mioglobina isso não fosse um problema, mas para proteínas mais sensíveis valores de pH muito diferentes podem causar mudanças conformacionais em suas estruturas. Foram realizadas reações nas temperaturas de 5 e 25 °C, e os mapas das restrições obtidas são apresentados na Figura 28, que já contém também os mapas da Figura 27 para facilitar a comparação com as reações em pH=6,5.



**Figura 28.** Mapa das ligações cruzadas identificadas para a mioglobina à 25 e 5 °C com ácido adípico como ALC em pH=7 (A) e pH=6,5 (B). Quando comparadas as mesmas temperaturas, observa-se a formação de um maior número de espécies com pH=6,5. Já quando duas reações no mesmo pH são comparadas, são observadas menos espécies nas reações em temperaturas mais baixas.

Analisando-se todos os mapas da figura, não se observa uma mudança muito significativa na cobertura geral da sequência da proteína ao se elevar o pH da reação ou alterar sua temperatura. Há, no entanto, algumas diferenças importantes a respeito do número e do tipo de espécies formadas nas quatro condições apresentadas. Como a visualização dessas diferenças através dos mapas não é tão direta, os dados advindos dessas reações foram organizados e são mostrados na Tabela 1.

**Tabela 1.** Número de restrições totais e divididas por espécie obtidas nos experimentos de ligação cruzada com ácido adípico em diferentes temperaturas e pH. É mostrado também a composição do número total da espécie de acordo com os resíduos reativos. Por exemplo, em pH 6,5 / 5 °C há 42 restrições, 18 do tipo XL e 24 do tipo ZL. Dentro das 18 espécies XL, 13 XL são entre lisinas (K-K), 5 entre lisina e serina (K-S) e nenhuma entre serinas (S-S).

	<i>Total</i>	<i>XL</i>	<i>K-K</i>	<i>K-S</i>	<i>S-S</i>	<i>ZL</i>	<i>K-D</i>	<i>K-E</i>	<i>S-D</i>	<i>S-E</i>
<b><i>pH 6,5 / 5 °C</i></b>	42	18	13	5	0	24	6	13	3	2
<b><i>pH 6,5 / 25 °C</i></b>	49	25	15	10	0	24	4	15	2	3
<b><i>pH 7 / 5 °C</i></b>	36	10	8	2	0	26	4	16	4	2
<b><i>pH 7 / 25 °C</i></b>	41	16	11	5	0	25	7	13	3	2

Pelos dados da tabela, nota-se que as tendências discutidas anteriormente para a variação de temperatura continuam a ser observadas, havendo uma diminuição número de espécies do tipo XL formadas devido ao desfavorecimento dos ataques nucleofílicos. Também se nota que a formação das espécies do tipo ZL não sofre grandes alterações com a com a redução da temperatura nas duas condições de pH.

Com a variação do pH reacional para 7, seria esperado um aumento no número de espécies XL formadas, já que mais resíduos de lisina estariam em sua forma desprotonada para efetuarem o ataque às espécies ativadas. No entanto, o comportamento oposto é que foi observado nas duas temperaturas. Uma possível explicação para esses resultados é que, ao aumentar o pH da reação, é aumentado também o número de íons  $\text{OH}^-$  do meio. A maior presença de hidroxilas aumenta a taxa de hidrólise das espécies ativadas com EDC e HOBt, já que podem atacar essas espécies e convertê-las novamente no ácido adípico. Como o número de choques das moléculas ativadas com o solvente é muito maior, menos espécies ativadas chegavam até a superfície da proteína, e menos espécies do tipo XL foram formadas.

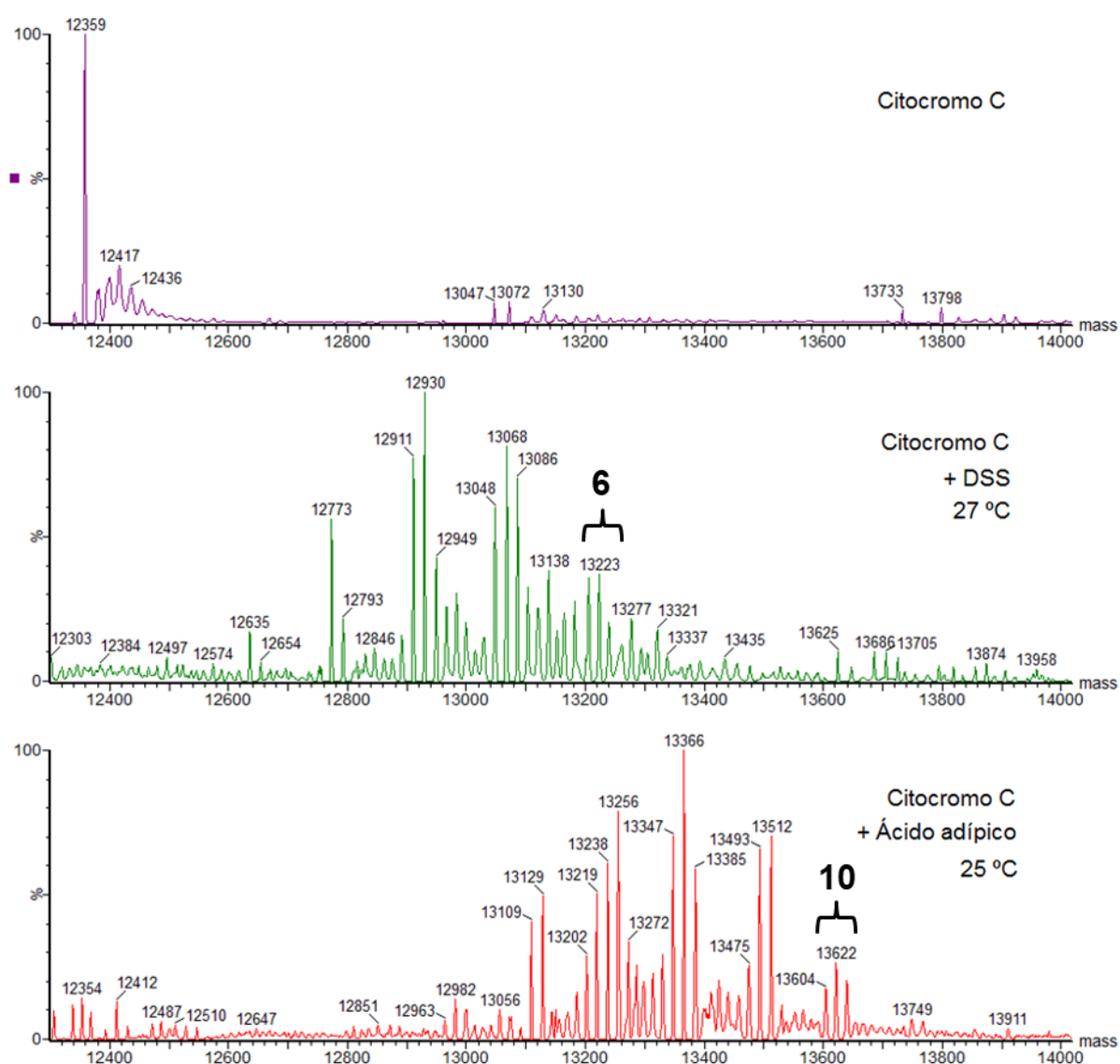
No caso das espécies ZL, seria esperada uma redução em sua formação com o aumento do pH pois, ao contrário das moléculas de ácido adípico que foram ativadas separadamente em pH=6,5, os resíduos ácidos só entraram em contato com EDC e HOBt em pH=7, e assim a etapa de ativação desses resíduos seria desfavorecida. Apesar disso, o que se observou foi que as quantidades formadas se mantêm praticamente constantes em todas as reações. Esse resultado foi surpreendente, e muito se especulou a respeito de uma explicação que fosse coerente com os dados. Buscando possíveis respostas que explicassem esses resultados, foi

encontrado na literatura um trabalho que avaliou a influência do pH nas reações de um ácido com o EDC<sup>65</sup>, correspondente à primeira reação da metodologia aqui proposta. Neste trabalho, foi observado que a queda na velocidade dessa reação na faixa de pH avaliada é muito pequena, e se torna relevante apenas a partir de pH=8. Dessa forma, a taxa de ativação dos resíduos de ácido aspártico e glutâmico com o EDC em pH=7 é praticamente igual à taxa de ativação do ácido adípico em pH=6,5, e a formação das espécies do tipo ZL se mantem constante nos dois valores de pH pois, devido aos resíduos estarem em contato, o fator hidrólise da espécie ativada não é relevante como na formação das espécies do tipo XL.

Para a continuidade do trabalho, foi mantida a condição de pH=6,5 para as reações de ligação cruzada, pois nela houve a formação de um maior número de espécies. Embora os testes com a variação do pH não permitiram controlar quais dos tipos de espécies são favorecidas na reação de ligação cruzada, nota-se que as quatro condições avaliadas nesses testes geraram mais espécies do que as reações utilizando o DSS.

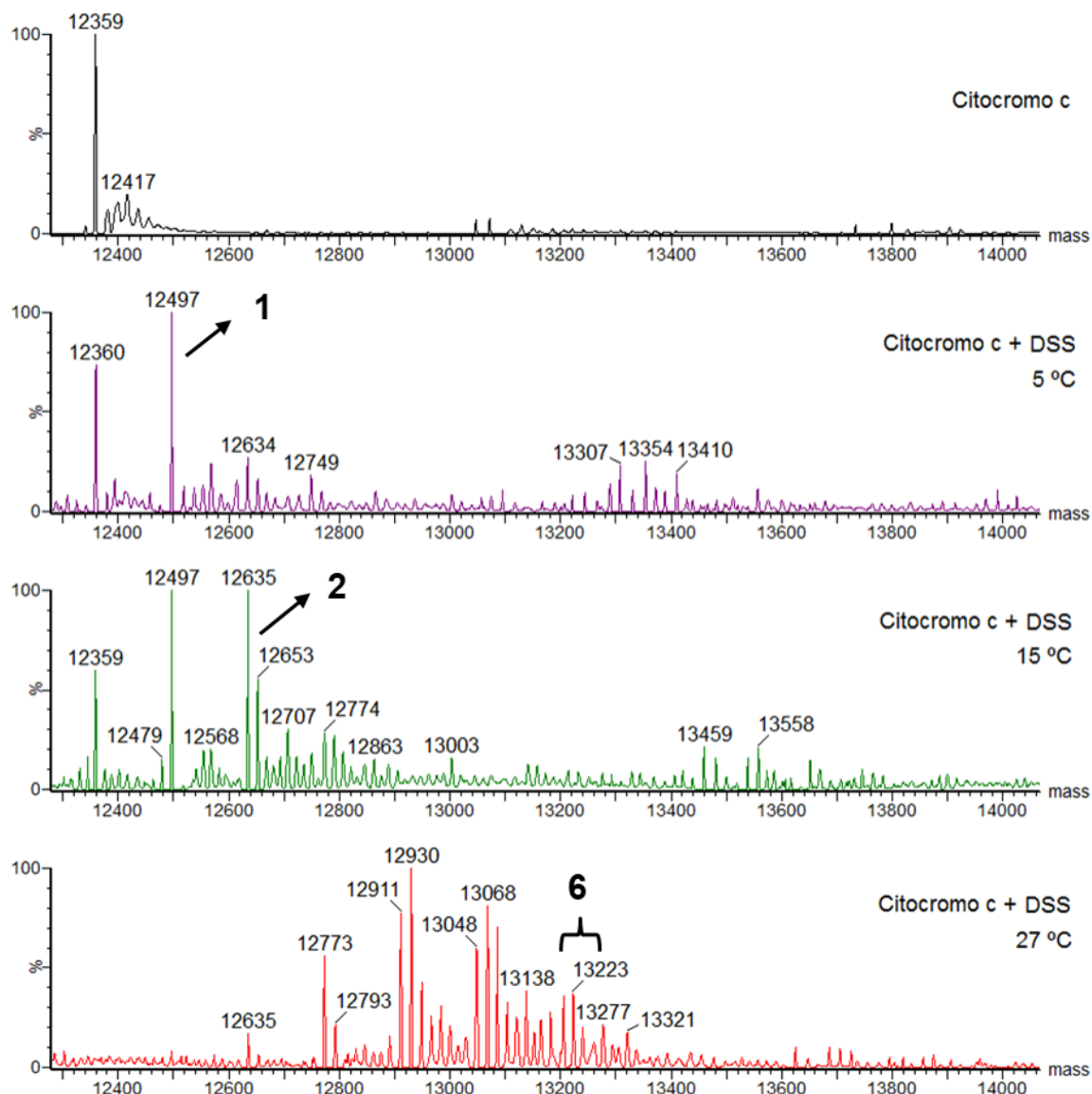
#### **4.5.Avaliação da metodologia proposta através da aplicação à proteína citocromo c**

Para avaliar se o método aqui proposto é robusto e apresenta as mesmas características para outras proteínas alvo, as reações comparativas entre DSS e ácido adípico foram refeitas, mas agora utilizando-se como proteína alvo o citocromo c. Foram utilizadas as condições otimizadas de excesso molar dos ALCs, período de ativação e tempo de reação, e foram testadas novamente as temperaturas de 5, 15 e 25 °C (27 °C para o DSS). Na Figura 29 são apresentados os espectros de MS das duas reações em suas temperaturas mais altas.



**Figura 29.** Espectros deconvoluídos de ESI(+)-MS de citocromo c comparando as reações de ligação cruzada utilizando o DSS e o ácido adípico como ALC. Mesmo na temperatura ótima da reação com DSS, sua taxa de incorporação por molécula da proteína é consideravelmente menor que a do ácido adípico.

Assim como para a mioglobina, a taxa de incorporação média do ácido adípico foi maior do que a do DSS, com 10 moléculas do ALC, chegando inclusive ao limite do número de incorporações considerados ideais, mostrando novamente uma maior eficiência desta metodologia. Na Figura 30 são apresentados os espectros de MS obtidos a partir das reações do DSS com o citocromo c nas diferentes temperaturas avaliadas.

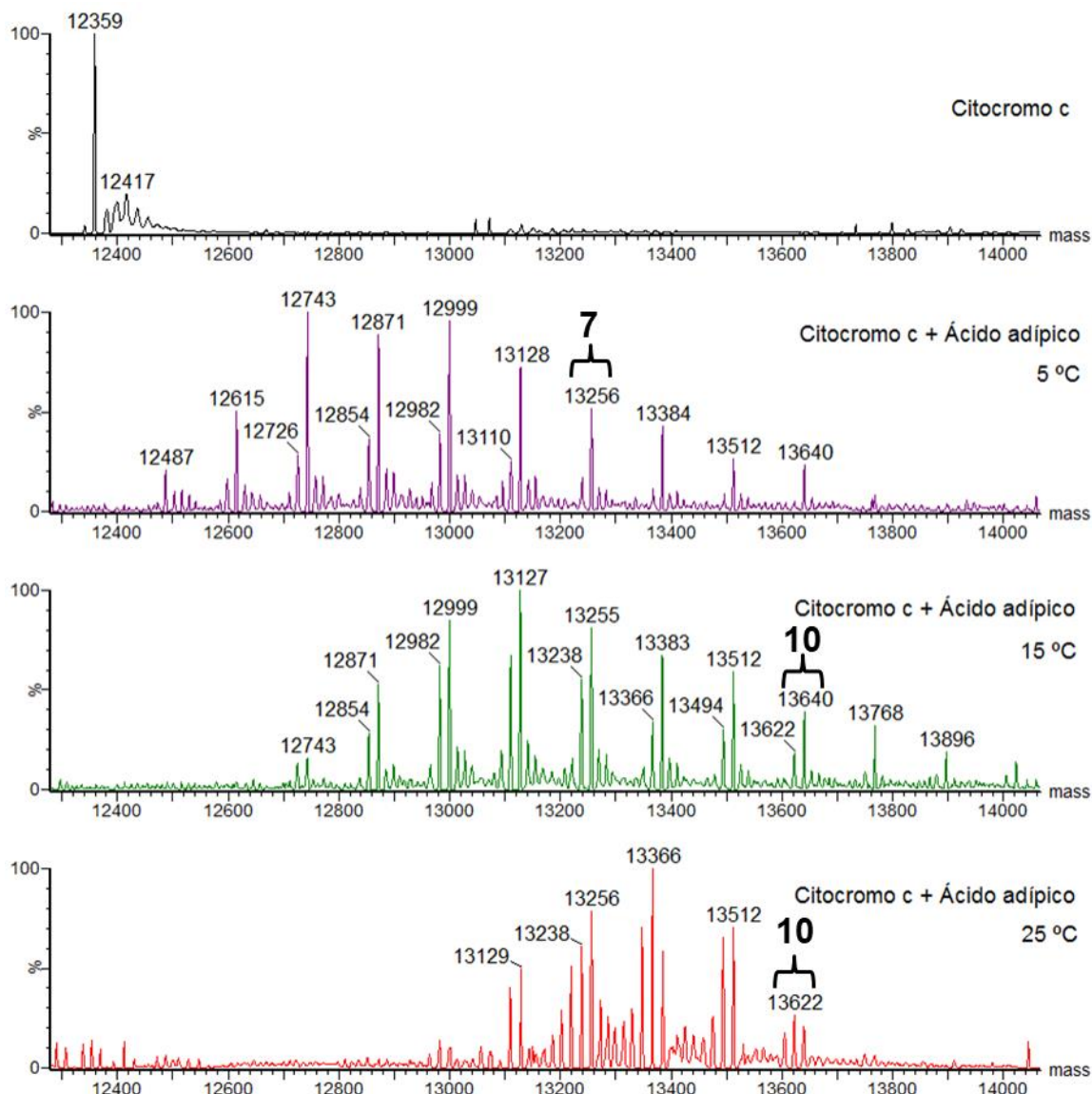


**Figura 30.** Espectros deconvoluídos de ESI(+)-MS de citocromo c nas reações de ligação cruzada para avaliação da influência da temperatura na taxa de incorporação de DSS. Observa-se que, com o decréscimo da temperatura, o rendimento da reação é muito reduzido.

Analisando-se a figura, novamente fica muito claro que o uso de DSS foi ineficaz em temperaturas mais baixas. Foram observadas as incorporações de apenas uma molécula do ALC na reação a 5 °C ( $m/z$  12497) e duas moléculas na reação a 15 °C ( $m/z$  12635), enquanto que a 27 °C foram observadas seis incorporações em média. Outro detalhe importante diz respeito a presença do sinal de  $m/z$  12359 nas condições de temperaturas mais baixas, indicando que a proteína não foi totalmente consumida pela reação.

Quando as mesmas condições foram aplicadas nas reações envolvendo o uso do ácido adípico, foram obtidos os espectros apresentados na Figura 31.





**Figura 31.** Espectros deconvoluídos de ESI(+)-MS de citocromo c nas reações de ligação cruzada para avaliação da influência da temperatura na taxa de incorporação de ácido adípico. Nota-se que, com o decréscimo da temperatura, a incorporação do ALC se mantém praticamente constante. A 5 °C, há a incorporação também de 10 espécies do ALC, mas apenas do tipo *dead-end*, sendo que com 7 incorporações, em média, há também a formação de espécies de ligação cruzada.

Pelos resultados mostrados nas Figura 30 e Figura 31, evidencia-se que a reação com o ácido adípico como ALC pode ser efetuada em temperaturas muito baixas. Entre as temperaturas de 25 e 15 °C a taxa de incorporação foi a mesma, com dez moléculas do ALC ( $m/z$  13640, correspondente a somente espécies do tipo *dead-end*). Na reação a 5 °C são observados os mesmos sinais característicos da incorporação de 10, 9 e 8 moléculas do ácido adípico. Porém, esses sinais

correspondem apenas à formação de espécies *dead-end*, sem a presença dos sinais de diferença -18 Da indicando a formação de espécies de XL. Por isso, para essa reação a incorporação média que levou à formação das espécies que fornecem restrições de distância entre os resíduos foi de 7 moléculas. Esse número de incorporações ainda é bem elevado, mostrando que de fato essa metodologia pode ser aplicada com sucesso em temperaturas mais baixas e, conseqüentemente, em uma gama maior de proteínas.

Comparando-se os resultados dos testes com a mioglobina e com o citocromo c, fica bastante claro o ganho dessa nova metodologia aqui proposta sobre a metodologia padrão de uso do DSS, visto que para ambas proteínas, mesmo as reações a 5 °C apresentaram taxa maior de incorporação do ácido adípico do que as reações utilizando o DSS em suas condições ideais, a 27 °C. Avaliando também os espectros de MS<sup>2</sup> da mioglobina, nota-se que com o uso do ácido adípico houve a formação de um número maior de espécies do tipo XL em relação ao DSS, mesmo em comparando-se as reações a 5 °C com as reações a 27 °C do DSS. Somando-se a isso, nessa metodologia há a formação conjunta de espécies do tipo ZL, em média dobrando o número de restrições obtidas em um único experimento de ligação cruzada e possibilitando uma cobertura maior da sequência da proteína, que deve auxiliar de forma significativa as etapas de modelagem.

Como consequência dos melhores rendimentos mesmo em baixas temperaturas, espera-se que a metodologia aqui proposta possa ser largamente empregada na análise estrutural de proteínas. Uma vasta gama de proteínas instáveis em solução em altas temperaturas, que não poderiam ser estudadas pelo uso de DSS e derivados de NHS em geral, agora são passíveis de serem analisadas através do uso do ácido adípico ativado com EDC e HOBt. Com isso, a metodologia deve expandir ainda mais a aplicabilidade da técnica de ligação cruzada no estudo de estruturas superiores de proteínas.

## 5. CONCLUSÕES

Neste projeto de mestrado foi desenvolvida uma nova metodologia para a técnica de ligação cruzada, utilizando como ALC o ácido adípico, duplamente ativado com EDC e HOBt. Os resultados obtidos mostram que esta metodologia possui várias vantagens em relação aos ésteres de NHS utilizados atualmente:

- 1) Maiores taxas de incorporação nas mesmas condições;
- 2) Menor tempo reacional;
- 3) Maior solubilidade do ALC;
- 4) Bons rendimentos mesmo em baixas temperaturas
- 5) Formação de produtos de *zero-length* conjuntamente aos de ligação cruzada convencional
- 6) Uso de reagentes mais acessíveis e estáveis em meio aquoso
- 7) Possibilidade de agregação de outras funcionalidades ao ALC.

Esta metodologia deve expandir ainda mais a aplicabilidade da técnica de ligação cruzada para a análise de estrutura de proteínas e complexos proteicos.

## 6. REFERÊNCIAS

1. Voet, D., Voet, J. G. & Pratt, C. W. *Fundamentals of Biochemistry: Life at the Molecular Level*. (John Wiley & Sons, Inc., 2013).
2. Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry*. (W.H. Freeman and Company, 2012). doi:10.1016/j.jse.2011.03.016
3. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).
4. Boehr, D. D. & Wright, P. E. BIOCHEMISTRY: How Do Proteins Interact? *Science (80-. )*. **320**, 1429–1430 (2008).
5. Robinson, C. V., Sali, A. & Baumeister, W. The molecular sociology of the cell. *Nature* **450**, 973–982 (2007).
6. Alber, F., Förster, F., Korkin, D., Topf, M. & Sali, A. Integrating Diverse Data for Structure Determination of Macromolecular Assemblies. *Annu. Rev. Biochem.* **77**, 443–477 (2008).
7. Sharon, M. How far can we go with structural mass spectrometry of protein complexes? *J. Am. Soc. Mass Spectrom.* **21**, 487–500 (2010).
8. Patwardhan, A. & Lawson, C. L. *Databases and Archiving for CryoEM. The Resolution Revolution: Recent Advances In cryoEM* **579**, (Elsevier Inc., 2016).
9. Doerr, A. Structural biology: Cryo-EM goes high-resolution. *Nat. Methods* **12**, 598–599 (2015).
10. Kendrew, J. C. *et al.* A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **181**, 662–666 (1958).
11. Perutz, M. F. *et al.* Structure of Hemoglobin: A Three-Dimensional Fourier Synthesis at 5.5Å. Resolution, Obtained by X-Ray Analysis. *Nature* **185**, 416–422 (1960).
12. Berman, H., Kleywegt, G., Nakamura, H. & Markley, J. The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future. *Structure* **20**, 391–396 (2012).
13. Wüthrich, K. The way to NMR structures of proteins. *Nat. Struct. Biol.* **8**, 923–925 (2001).
14. RCSB Protein Data Bank - PDB. Available at: <http://www.rcsb.org/pdb/home/home.do>. (Accessed: 11th June 2017)
15. Sali, A., Glaeser, R., Earnest, T. & Baumeister, W. From words to literature in

- structural proteomics. *Nature* **422**, 216–225 (2003).
16. Ward, A. B., Sali, A. & Wilson, I. A. Integrative Structural Biology. *Science*, **339**, 913–915 (2013).
  17. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **10**, 980–980 (2003).
  18. Barker, J., Ando, D. J. & Davis, R. *Mass spectrometry: Analytical Chemistry by Open Learning*. (John Wiley & Sons, 1999).
  19. Fenn, J., Mann, M., Meng, C., Wong, S. & Whitehouse, C. Electrospray ionization for mass spectrometry of large biomolecules. *Science* (80-. ). **246**, 64–71 (1989).
  20. Karas, M. & Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60**, 2299–2301 (1988).
  21. Kinter, M. & Sherman, N. E. *Protein Sequencing and Identification Using Tandem Mass Spectrometry. Wiley-Interscience Series on Mass Spectrometry* (John Wiley & Sons, Inc., 2000). doi:10.1002/0471721980
  22. Hu, Q. *et al.* The Orbitrap: A new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443 (2005).
  23. Michalski, A. *et al.* Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol. Cell. Proteomics* **10**, M111.011015-M111.011015 (2011).
  24. Aebersold, R. & Goodlett, D. R. Mass Spectrometry in Proteomics. *Chem. Rev.* **101**, 269–296 (2001).
  25. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
  26. Patterson, S. D. & Aebersold, R. H. Proteomics: the first decade and beyond. *Nat. Genet.* **33**, 311–323 (2003).
  27. Steen, H. & Mann, M. The abc's (and xyz's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **5**, 699–711 (2004).
  28. Yates, J. R. Mass Spectral Analysis in Proteomics. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 297–316 (2004).
  29. Domon, B. & Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* (80-. ). **312**, 212–217 (2006).
  30. Cravatt, B. F., Simon, G. M. & Yates III, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* **450**, 991–1000 (2007).

31. Bai, Y. W., Milne, J. S., Mayne, L. & Englander, S. W. Primary Structure Effects on Peptide Group Hydrogen-Exchange. *Proteins Struct. Funct. Genet.* **17**, 75–86 (1993).
32. Xu, G. & Chance, M. R. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **107**, 3514–3543 (2007).
33. Maurer, M. M., Donohoe, G. C. & Valentine, S. J. Advances in ion mobility-mass spectrometry instrumentation and techniques for characterizing structural heterogeneity. *Analyst* **140**, 6782–6798 (2015).
34. Sinz, A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.* **38**, 1225–1237 (2003).
35. Wong, S. S. *Chemistry of protein conjugation and cross-linking*. (CRC Press, 1991).
36. Young, M. M. *et al.* High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci.* **97**, 5802–5806 (2000).
37. Holding, A. N. XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods* **89**, 54–63 (2015).
38. Santos, A. M. *et al.* FERM domain interaction with myosin negatively regulates FAK in cardiomyocyte hypertrophy. *Nat. Chem. Biol.* **8**, 102–110 (2011).
39. Pereira, M. B. M. *et al.*  $\alpha$ B-crystallin interacts with and prevents stress-activated proteolysis of focal adhesion kinase by calpain in cardiomyocytes. *Nat. Commun.* **5**, 5159 (2014).
40. Chen, Z. A. *et al.* Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* **29**, 717–726 (2010).
41. Carven, G. J. & Stern, L. J. Probing the ligand-induced conformational change in HLA-DR1 by selective chemical modification and mass spectrometric mapping. *Biochemistry* **44**, 13625–13637 (2005).
42. Huang, B. X. Interdomain Conformational Changes in Akt Activation Revealed by Chemical Cross-linking and Tandem Mass Spectrometry. *Mol. Cell. Proteomics* **5**, 1045–1053 (2006).
43. Huang, B. X. & Kim, H.-Y. Probing Akt-inhibitor interaction by chemical cross-linking and mass spectrometry. *J. Am. Soc. Mass Spectrom.* **20**, 1504–1513 (2009).

44. Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein–protein interactions. *Mass Spectrom. Rev.* **25**, 663–682 (2006).
45. Andrews, G. L., Simons, B. L., Young, J. B., Hawkrigde, A. M. & Muddiman, D. C. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal Chem* **83**, 5442–5446 (2011).
46. Paulo, J. A. *et al.* Quantitative mass spectrometry-based multiplexing compares the abundance of 5000 *S. cerevisiae* proteins across 10 carbon sources. *J. Proteomics* **148**, 85–93 (2016).
47. Lima, D. B. *et al.* SIM-XL: A powerful and user-friendly tool for peptide cross-linking analysis. *J. Proteomics* **129**, 51–55 (2015).
48. Belsare, K. D. *et al.* P-LinK: A method for generating multicomponent cytochrome P450 fusions with variable linker length. *Biotechniques* **57**, 13–20 (2014).
49. Sarpe, V. *et al.* High Sensitivity Crosslink Detection Coupled With Integrative Structure Modeling in the Mass Spec Studio. *Mol. Cell. Proteomics* **15**, 3071–3080 (2016).
50. Götze, M. *et al.* StavroX-A software for analyzing crosslinked products in protein interaction studies. *J. Am. Soc. Mass Spectrom.* **23**, 76–87 (2012).
51. Chu, F., Mahrus, S., Craik, C. S. & Burlingame, A. L. Isotope-coded and affinity-tagged cross-linking (ICATXL): An efficient strategy to probe protein interaction surfaces. *J. Am. Chem. Soc.* **128**, 10362–10363 (2006).
52. Kao, A. *et al.* Development of a Novel Cross-linking Strategy for Fast and Accurate Identification of Cross-linked Peptides of Protein Complexes. *Mol. Cell. Proteomics* **10**, M110.002212-M110.002212 (2011).
53. Müller, M. Q., Dreiocker, F., Ihling, C. H., Schäfer, M. & Sinz, A. Cleavable Cross-Linker for Protein Structure Analysis: Reliable Identification of Cross-Linking Products by Tandem MS. *Anal. Chem.* **82**, 6958–6968 (2010).
54. Petrotchenko, E. V., Serpa, J. J. & Borchers, C. H. An Isotopically Coded CID-cleavable Biotinylated Cross-linker for Structural Proteomics. *Mol. Cell. Proteomics* **10**, M110.001420-M110.001420 (2011).
55. Krauth, F., Ihling, C. H., Rüttinger, H. H. & Sinz, A. Heterobifunctional isotope-labeled amine-reactive photo-cross-linker for structural investigation of proteins by matrix-assisted laser desorption/ionization tandem time-of-flight and

- electrospray ionization LTQ-Orbitrap mass spectrometry. *Rapid Commun. Mass Spectrom.* **23**, 2811–2818 (2009).
56. Gomes, A. F. & Gozzo, F. C. Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI- and ESI-MS/MS. *J. Mass Spectrom.* **45**, 892–899 (2010).
  57. Trnka, M. J., Baker, P. R., Robinson, P. J. J., Burlingame, A. L. & Chalkley, R. J. Matching Cross-linked Peptide Spectra: Only as Good as the Worse Identification. *Mol. Cell. Proteomics* **13**, 420–434 (2014).
  58. Solomons, T. W. G. & Fryhle, C. B. *Organic Chemistry*. (John Wiley & Sons, Inc., 2011).
  59. Rivera-Santiago, R. F., Sriswasdi, S., Harper, S. L. & Speicher, D. W. Probing structures of large protein complexes using zero-length cross-linking. *Methods* **89**, 99–111 (2015).
  60. Leitner, A. *et al.* Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci.* **111**, 9455–9460 (2014).
  61. Clayden, J., Greeves, N. & Warren, S. *Organic Chemistry*. (Oxford University Press, 2012).
  62. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* **27**, (1999).
  63. TOPOLINK - Institute of Chemistry - UNICAMP. Available at: <http://leandro.iqm.unicamp.br/topolink/usage.shtml>. (Accessed: 18th June 2017)
  64. Atkins, P. W. & Paula, J. de. *Physical Chemistry*. (W. H. Freeman and Company, 2010).
  65. Chan, L. C. & Cox, B. G. Kinetics of Amide Formation through Carbodiimide/ N-Hydroxybenzotriazole (HOBt) Couplings. (2007). doi:10.1021/jo701558y